Heinemann

PHYSICS 11

4TH EDITION

VCE Units 1 & 2
Written for the VCE Physics
Study Design 2016–2021

Greg Moran
Doug Bail
Keith Burrows
Rob Chapman
Ann Conibear
Carmel Fry
David Madden
John Nicholson
Jeffrey Stanger
Daniela Nardelli
Paul Cuthbert
Mark Hamilton
Alistair Harkness
Jack Jurica
Craig Tilley

ALWAYS LEARNING

PEARSON

Heinemann

PHYSICS 11

4TH EDITION

VCE Units 1 & 2

COORDINATING AUTHOR:
Greg Moran

AUTHORS:
Doug Bail
Keith Burrows
Rob Chapman
Ann Conibear
Carmel Fry
David Madden
John Nicholson
Jeffrey Stanger

CONTRIBUTING AUTHORS:
Daniela Nardelli
Paul Cuthbert
Mark Hamilton
Alistair Harkness
Jack Jurica
Craig Tilley

Heinemann
PHYSICS 11
4TH EDITION

Writing and Development Team

We are grateful to the following people for their time and expertise in contributing to the Heinemann Physics 11 project.

Malcolm Parsons
Publishing Consultant
Series Consultant

Greg Moran
Head of Department
Coordinating Author

Daniela Nardelli
Learning Development Leader
Subject Lead and Author (Print)

Craig Tilley
Science Writer and Assessor
Associate Subject Lead (Print)

Doug Bail
Education Consultant
Author

Rachael Barker
Teacher
Reviewer

Keith Burrows
Educator
Author

Rob Chapman
Educator
Author

Ann Conibear
Teacher
Author

Paul Cuthbert
Teacher
Author

Carmel Fry
Head of Senior Years Curriculum
Author

Mark Hamilton
Teacher
Author

Alistair Harkness
Teacher
Author

Matt Jones
Teacher
Reviewer

Jack Jurica
Teacher
Author

Melanie Lane
Head of Science
Author

David Madden
Acting Manager, QCAA
Author

Svetlana Marchouba
Laboratory Technician
Author

Dr John Nicholson
Head of Department
Author

Jeffrey Stanger
Head of Department
Author

Contents

Unit 1: What ideas explain the physical world?

AREA OF STUDY 1

How can thermal effects be explained?

AREA OF STUDY 2

How do electric circuits work?

AREA OF STUDY 3

What is matter and how is it formed

Unit 2: What do experiments reveal about the physical world?

AREA OF STUDY 1

How can motion be described and explained?

AREA OF STUDY 2
Options

AREA OF STUDY 3
Practical investigation

How to use this book

Heinemann Physics 11 4th edition

Heinemann Physics 11 4th edition has been written to the new VCE Physics Study Design 2016 – 2021. The book covers Units 1 and 2 in an easy-to-use resource. Explore how to use this book below.


CHAPTER 01 Heating processes

Key knowledge

Chapter opener

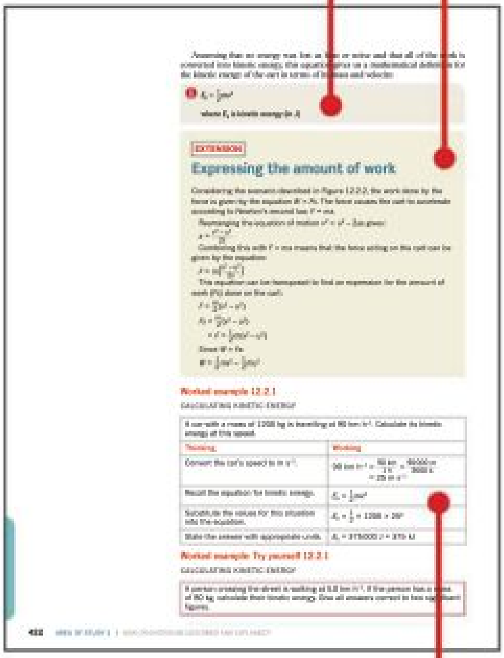
Chapter opening pages links the Study Design to the chapter content. Key knowledge addressed in the chapter is clearly listed.

Extension

Extension material goes beyond the core content of the Study Design. It is intended for students who wish to expand their depth of understanding.

Highlight

Focus on important information such as key definitions, formulae and summary points.


EXTENSION

Expressing the amount of work

Worked example 12.2.1

CALCULATING KINETIC ENERGY

Thinking

Working

Worked example: Try yourself 12.2.1

CALCULATING KINETIC ENERGY


PHYSICS IN ACTION

Wind chill

PHYSICSFILE

Paragliders

Physics in Action

Physics in Action place physics in an applied situation or relevant context. These refer to the nature and practice of physics, applications of physics and the associated issues and the historical development of concepts and ideas.

PhysicsFile

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

Worked examples

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

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

Fully worked solutions to all Try Yourself: Worked examples are available on *Heinemann Physics 11 4th edition ProductLink*.

Chapter review

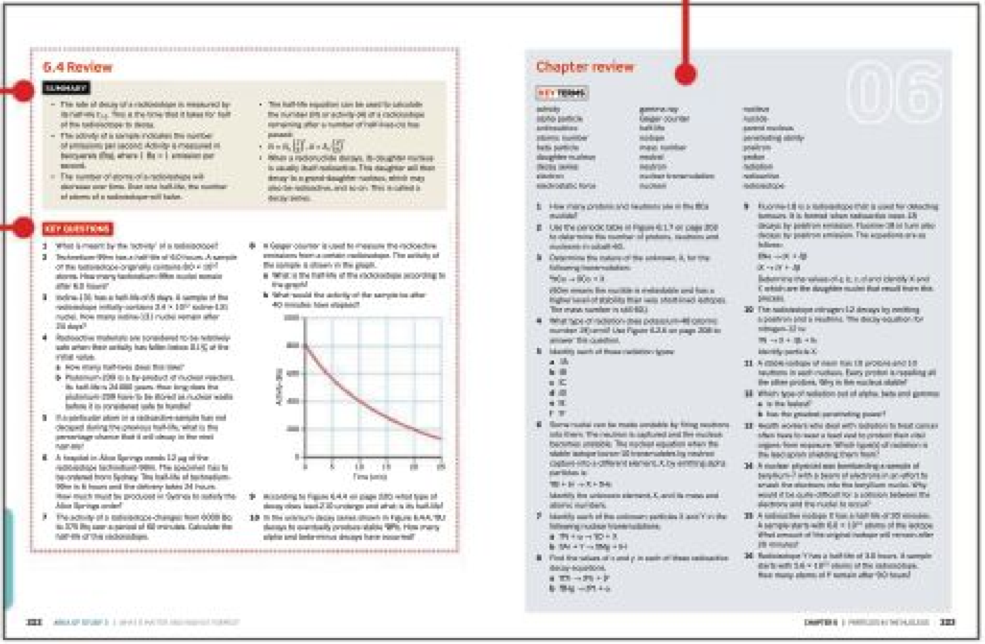
Each chapter finishes with a set of higher order questions to test students' ability to apply the knowledge gained from the chapter.

Section summary

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

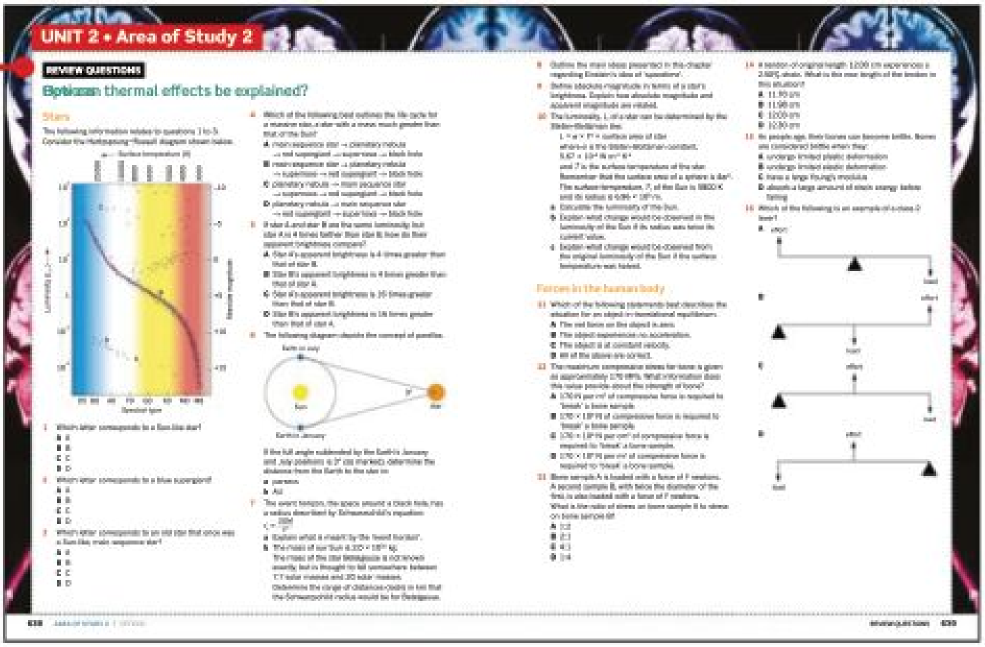
Section review

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



Area of Study review

Each Area of Study finishes with a comprehensive set of exam-style questions, including multiple choice and extended response, that assist students draw together their knowledge and understanding and apply it to this style of questions.



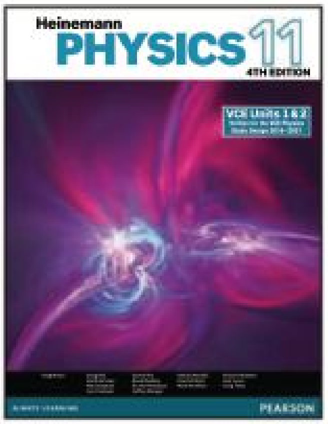
Answers

Numerical answers and key short response answers are included at the back of the book. Comprehensive answers and fully worked solutions for all section review questions, Try Yourself: Worked examples, chapter review questions and Area of Study review questions are provided via *Heinemann Physics 11 4th edition ProductLink.*

Glossary

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

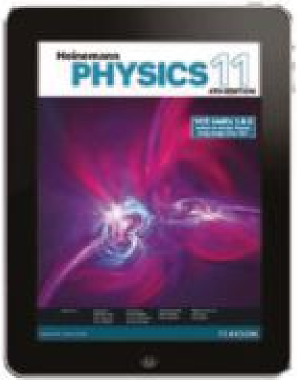
Heinemann Physics 11
4th edition



Student Book

Heinemann Physics 11 4th Edition has been written to fully align with the *2016–2021 VCE Physics Study Design*.

All core components of Units 1 and 2 are included in the book. Unit 2 Area of Study 2 includes six Option Chapters available on Pearson eBook 3.0. Students undertake one option.



Pearson eBook 3.0

Pearson eBook 3.0 is the digital version of the student book. It retains the integrity of the printed page and is available online or offline on any device – PC/Mac, Android tablet, iPad.

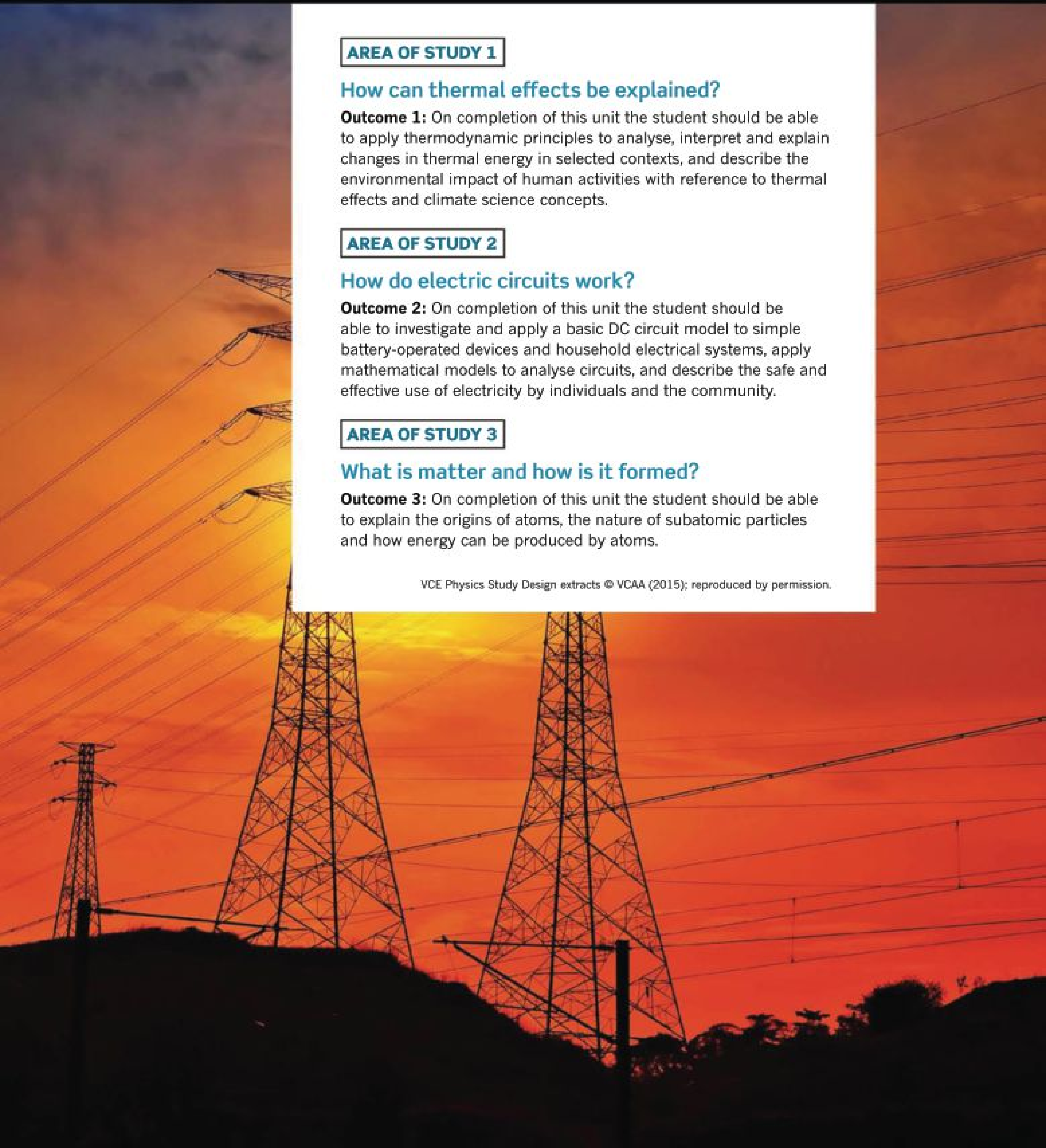
Students can access the six Unit 2 Area of Study 2 Options Chapters via the Pearson eBook 3.0.

ProductLink

Heinemann Physics 11 4th Edition ProductLink provides comprehensive support for teachers and students completely free of charge. ProductLink includes comprehensive answers and worked solutions to all student book questions, teaching programs, risk assessments and more.

UNIT 1

What ideas explain the physical world?



AREA OF STUDY 1

How can thermal effects be explained?

Outcome 1: On completion of this unit the student should be able to apply thermodynamic principles to analyse, interpret and explain changes in thermal energy in selected contexts, and describe the environmental impact of human activities with reference to thermal effects and climate science concepts.

AREA OF STUDY 2

How do electric circuits work?

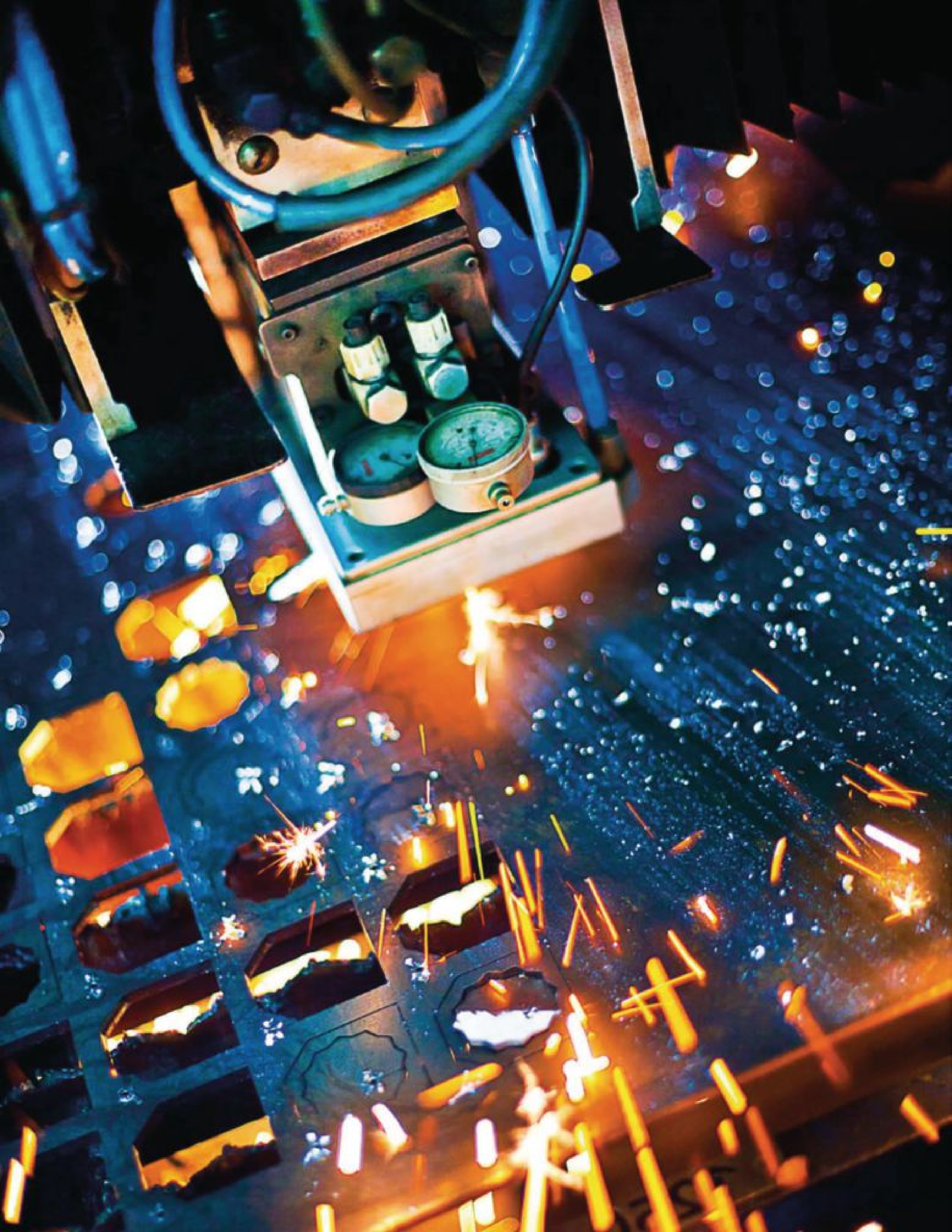
Outcome 2: On completion of this unit the student should be able to investigate and apply a basic DC circuit model to simple battery-operated devices and household electrical systems, apply mathematical models to analyse circuits, and describe the safe and effective use of electricity by individuals and the community.

AREA OF STUDY 3

What is matter and how is it formed?

Outcome 3: On completion of this unit the student should be able to explain the origins of atoms, the nature of subatomic particles and how energy can be produced by atoms.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.



CHAPTER 01 Heating processes

Due to increasing levels of carbon dioxide in the atmosphere, the Earth is getting warmer. The last two decades of the twentieth century were the warmest for over 400 years. In 2014, the Earth was the warmest since records began in 1890. These higher temperatures increase the severity of bushfires. The rate of evaporation from pastures also increases, drying the land and reducing the level of food production.

Thermal energy is part of our everyday experience. Humans can thrive in the climatic extremes of the Earth, from the outback deserts to ski slopes in winter.

Key knowledge

By the end of this chapter, you will have covered material from the study of the thermodynamic principles related to heating processes, including concepts of temperature, energy and work, and will be able to:

- convert temperature between degrees Celsius and kelvin
- describe the zeroth law of thermodynamics as two bodies in contact with each other coming to a thermal equilibrium
- describe temperature with reference to the average kinetic energy of the atoms and molecules within a system
- investigate and apply theoretically and practically the first law of thermodynamics to simple situations: $Q = U + W$
- explain internal energy as the energy associated with random disordered motion of molecules
- distinguish between conduction, convection and radiation with reference to heat transfer within and between systems
- investigate and analyse theoretically and practically the energy required to:
  - raise the temperature of a substance: $Q = mc\Delta T$
  - Change the state of a substance: $Q = mL$
- Explain why cooling results from evaporation using a simple kinetic energy model.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

1.1 Heat and temperature

In the sixteenth century, Sir Francis Bacon, an English essayist and philosopher, proposed a radical idea: that heat is motion. He went on to write that heat is the rapid vibration of tiny particles within every substance. At the time, his ideas were dismissed because the nature of particles wasn't fully understood. An opposing theory at the time was that heat was related to the movement of a fluid called 'caloric' that filled the spaces within a substance.

Today, it is understood that all matter is made up of small particles (atoms or molecules). Using this knowledge, it is possible to look more closely at what happens during heating processes.

This section starts by looking at the kinetic particle model, which states that the small particles (atoms or molecules) that make up all matter have kinetic energy. This means that all particles are in constant motion, even in extremely cold solids. It was thought centuries ago that if a material was continually made cooler, there would be a point at which the particles would eventually stop moving. This coldest possible temperature is called **absolute zero** and will be discussed later in this section.

KINETIC PARTICLE MODEL

Some philosophers of the Middle Ages believed that heat was a fluid that filled the spaces between the particles of a substance and flowed from one substance to another. This is known as the 'caloric' theory. When caloric flowed from one substance into another, the first object cooled down and the second object heated up. Many attempts were made to detect caloric, but none were successful. It was assumed that caloric had no mass, odour, taste or colour. Scientists now know that caloric simply doesn't exist.

Much of our understanding of the behaviour of matter today depends on a model called the **kinetic particle model** (kinetic theory).

Recall that a model is a representation that describes or explains the workings within an object, system or idea.

These are the assumptions behind the kinetic particle model:

- All matter is made up of many very small particles (atoms or molecules).
- The particles are in constant motion.
- No kinetic energy is lost or gained overall during collisions between particles.
- There are forces of attraction and repulsion between the particles in a material.
- The distances between particles in a gas are large compared with the size of the particles.

The kinetic theory applies to all states (or phases) of matter: solids, liquids and gases.

Solids

Within a solid, the particles must be exerting attractive forces or bonds on each other for the matter to hold together in its fixed shape. There must also be repulsive forces, without which the attractive forces would cause the solid to collapse. In a solid, the attractive and repulsive forces hold these particles in more or less fixed positions, usually in a regular arrangement or lattice (see Figure 1.1.1(a)). But the particles in a solid are not completely still; they vibrate around average positions. The forces on individual particles are sometimes predominantly attractive and sometimes repulsive, depending on their exact position relative to neighbouring particles.

Liquids

Within a liquid, there is still a balance of attractive and repulsive forces. Compared with a solid, the particles in a liquid have more freedom to move around each other and will therefore take the shape of the container (see Figure 1.1.1(b)). Generally, the liquid takes up a slightly greater volume than it would in the solid state. Particles collide but remain attracted to each other, so the liquid remains within a fixed volume but with no fixed shape.

Gases

In a gas, particles are in constant, random motion, colliding with each other and the walls of the container. The particles move rapidly in every direction, quickly filling the volume of any container and occasionally colliding with each other (see Figure 1.1.1(c)). A gas has no fixed volume. The particle speeds are high enough that, when the particles collide, the attractive forces are not strong enough to keep the particles close together. The repulsive forces cause the particles to separate and move off in other directions.

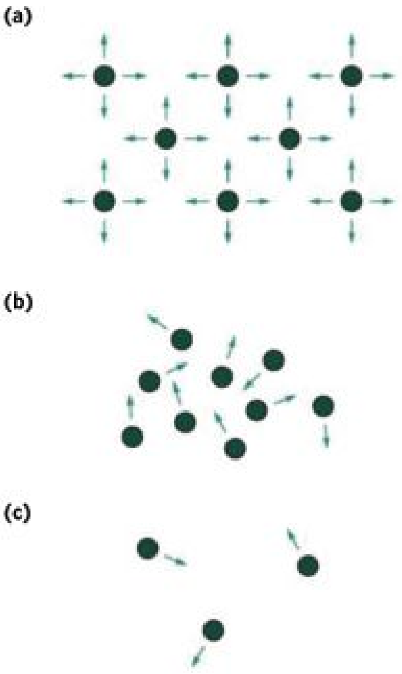


FIGURE 1.1.1 (a) Molecules in a solid have low kinetic energy and vibrate around average positions within a regular arrangement. (b) The particles in a liquid have more kinetic energy than those in a solid. They move more freely and will take the shape of the container. (c) Gas molecules are free to move in any direction.

PHYSICSFILE

States of matter: Plasma

The three basic phases (states) of matter are solid, liquid and gas. These are generally all that are discussed in secondary science.

There are in fact four phases of matter that are observable in everyday life—solid, liquid, gas and plasma. Under special conditions, several more exist. Plasma exists when matter is heated to very high temperatures and electrons are freed (ionisation). A gas that is ionised and has an equal number of positive and negative charges is called plasma. The interior of stars consists of plasma. In fact, most of the matter in the universe is plasma (Figure 1.1.2).

FIGURE 1.1.2 99.9 per cent of the visible universe is made up of plasma.

THE KINETIC PARTICLE MODEL, INTERNAL ENERGY AND TEMPERATURE

The kinetic particle model can be used to explain the idea of heat as a transfer of energy. **Heat** (measured in joules) is the transfer of **thermal energy** from a hotter body to a colder one. Heating is observed by the change in **temperature**, the change of state or the expansion of a substance.

When a solid substance is 'heated', the particles within the material gain either **kinetic energy** (move faster) or **potential energy** (move away from their equilibrium positions).

The term heat refers to energy that is being transferred (moved). So it is incorrect to talk about heat contained in a substance. The term **internal energy** refers to the total kinetic and potential energy of the *particles within* a substance. Heating (the transfer of thermal energy) changes the internal energy of a substance by affecting the kinetic energy and/or potential energy of the particles within the substance. The movement of the particles in a substance due to kinetic energy is ordered: the particles move back and forth and we can model their behaviour. In comparison, the internal energy of a system is associated with the chaotic motion of the particles—it concerns the behaviour of a large number of particles that all have their own kinetic and potential energy.

PHYSICS IN ACTION

Energy

Energy is a very important concept in the study of the physical world, and is a focus in all areas of scientific study. Later chapters investigate energy in more detail.

Energy is a measure of an object's ability to do **work**. For example, raising an object's temperature or lifting an object is referred to as doing work. Work is measured in joules. The symbol for joules is J. Figure 1.1.3 shows the amount of joules available from some energy sources.

Kinetic energy is the energy of movement. It is equal to the amount of work needed to bring an object from rest to its present speed or to return it to rest. Potential energy is stored energy. There are many forms of potential energy, for example gravitational, nuclear, spring and chemical. Chemical potential energy is associated with the bonds between the particles of a substance. An increase in the potential energy of particles in a substance results in movement of the particles from their equilibrium positions.

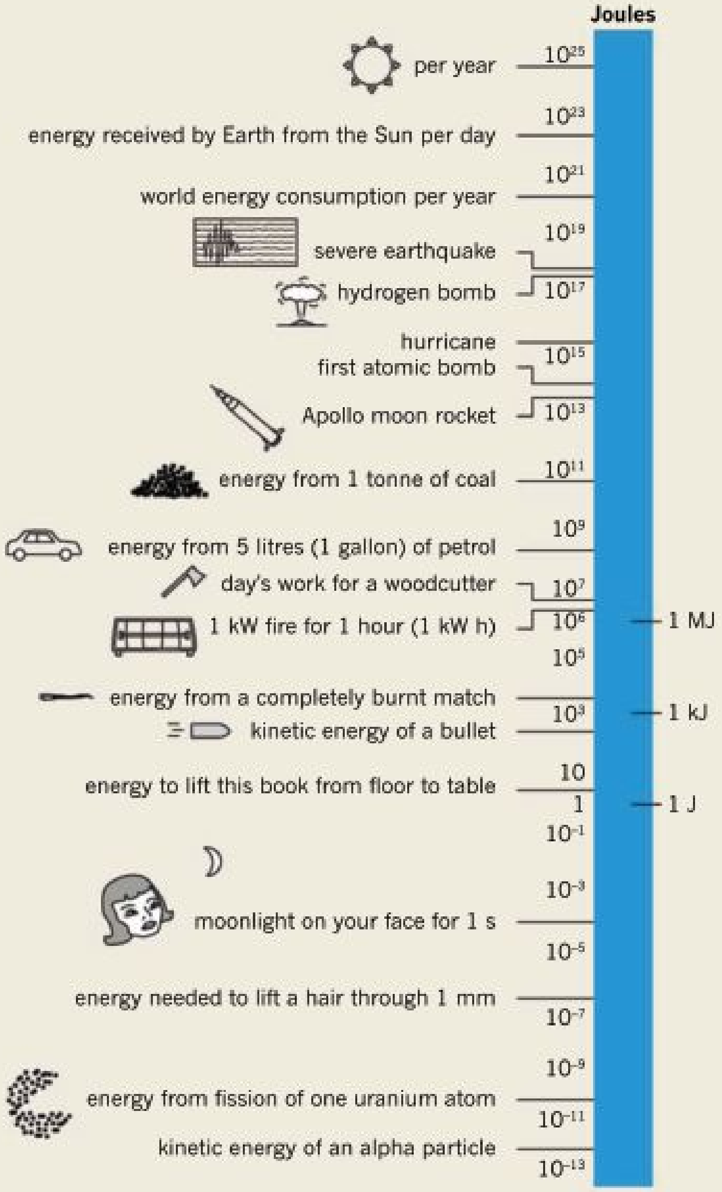


FIGURE 1.1.3 The comparative amounts of energy available from several sources.

- Heating is a process that always transfers thermal energy from a hotter substance to a colder substance.
- Heat is measured in joules (J).
- Temperature is related to the average kinetic energy of the particles in the substance. The faster the particles move, the higher the temperature of the substance.

Using the kinetic particle model, an increase in the total internal energy of the particles in a substance will result in an increase in temperature if there is a net gain in kinetic energy. Hot air balloons are an example of this process in action. The air in a hot air balloon is heated by a gas burner to a maximum of 120°C. The nitrogen (78%) and oxygen (21%) molecules in the hot air gain energy and so move a lot faster. The air in the balloon becomes less dense than the surrounding air, causing the balloon to float as seen in Figure 1.1.4.

Sometimes heating results only in the change of state or expansion of an object, and not a change in temperature. In these cases, the total internal energy of the particles has still increased but only the potential energy has increased, not the kinetic energy.

For instance, particles in a solid being heated will continue to be mostly held in place, due to the relatively strong interparticle forces. For the substance to change state from solid to liquid, it must receive enough energy to separate the particles from each other and disrupt the regular arrangement of the solid. During this 'phase change' process, the energy is used to overcome the strong interparticle forces, but not to change the overall speed of the particles. In this situation, the temperature does not change. This will be discussed in more detail in Section 1.3 'Latent heat'.

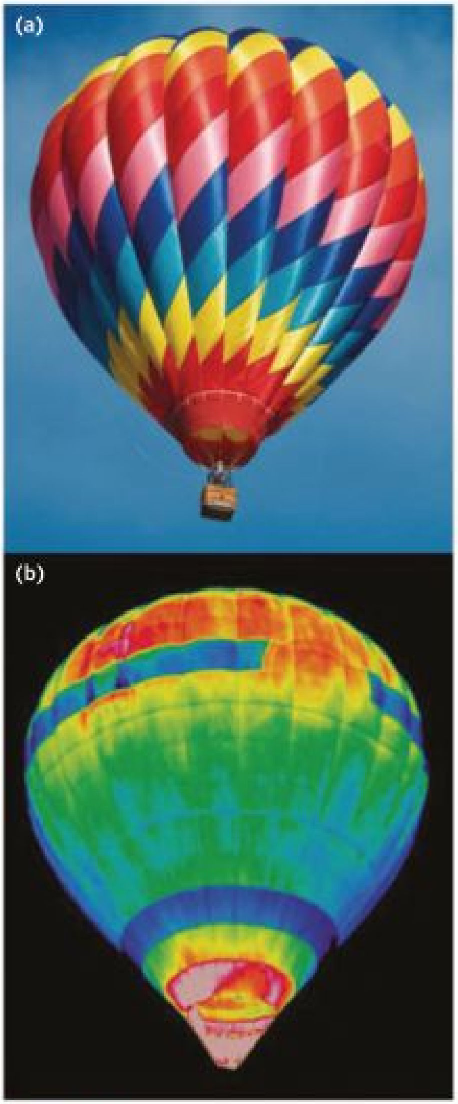


FIGURE 1.1.4 (a) Nitrogen and oxygen molecules gain energy when the air is heated, lowering the density of the air and causing the hot air balloon to rise off the ground. (b) A thermal image shows the temperature of the air inside the balloon.

MEASURING TEMPERATURE

Only four centuries ago, there were no thermometers and people described heating effects by vague terms such as hot, cold and lukewarm. In about 1593, Italian inventor Galileo Galilei made one of the first thermometers. His 'thermoscope' was not particularly accurate as it did not take into account changes in air pressure, but it did suggest some basic principles for determining a suitable scale of measurement. His work suggested that there be two fixed points: the hottest day of summer and the coldest day of winter. A scale like this is referred to as an arbitrary scale, because the fixed points are randomly chosen.

Celsius and Fahrenheit scales

Two of the better known arbitrary temperature scales are the Fahrenheit and Celsius scales. Gabriel Fahrenheit of Germany invented the first mercury thermometer in 1714. While Fahrenheit is used in the United States of America to measure temperature, the system used in most of the countries of the world is the Celsius scale.

Absolute scales are different from arbitrary scales. For a scale to be regarded as 'absolute', it should have no negative values. The fixed points must be reproducible and have zero as the lowest value.

Kelvin scale

When developing the absolute temperature scale, the triple point of water provides one reliable fixed point. This is a point where the combination of temperature and air pressure allows all three states of water to coexist. For water, the triple point is only slightly above the standard freezing point (0.01°C) and provides a unique and repeatable temperature with which to adjust the Celsius scale.

The absolute or **kelvin** temperature scale is based on absolute zero and the triple point of water. See Figure 1.1.5 for a comparison of the kelvin and Celsius scales.

- The freezing point of water (0°C) is equivalent to 273.15 K (kelvin). This is often approximated to 273 K.
- The size of each unit, 1°C or 1 K, is the same.
- The word 'degree' and the degree symbol are not used with the kelvin scale.
- To convert a temperature from degrees Celsius to kelvin: add 273.
- To convert a temperature from kelvin to degrees Celsius: subtract 273.

- 0 degrees Celsius (0°C) is the freezing point of water at standard atmospheric pressure.
- 100°C is the boiling point of water at standard atmospheric pressure.

PHYSICSFILE

Unit conventions in physics

The unit for energy, the joule, is named after James Joule in recognition of his work. When a unit is named after a person, its symbol is usually a capital letter but the unit name is always lower case, e.g. joule (J), newton (N), kelvin (K).

Exceptions are degrees Celsius (°C) and degrees Fahrenheit (°F), which also include a degree symbol.

Units not named after people usually have both the symbol and the name in lowercase, e.g. metre (m), second (s).

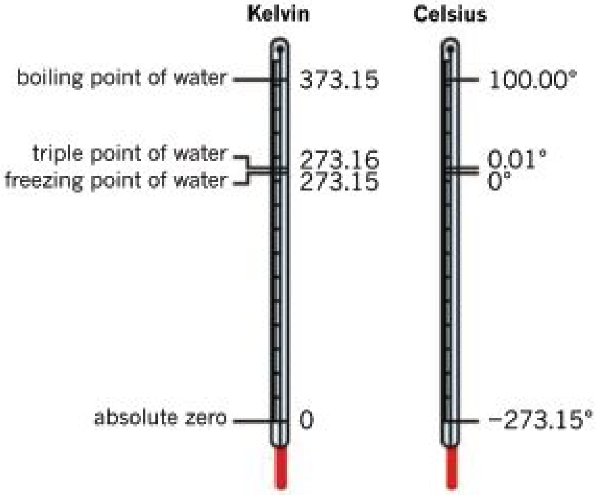


FIGURE 1.1.5 Comparison of the kelvin and Celsius scales. Note that there are no negative values on the kelvin scale.

- Absolute zero = 0 K = −273.15°C
- All molecular motion ceases at absolute zero. This is the coldest temperature possible.

Absolute zero

Experiments indicate that there is a limit to how cold things can get. The kinetic theory suggests that if a given quantity of gas is cooled, its volume decreases. The volume can be plotted against temperature and results in a straight line graph as shown in Figure 1.1.6.

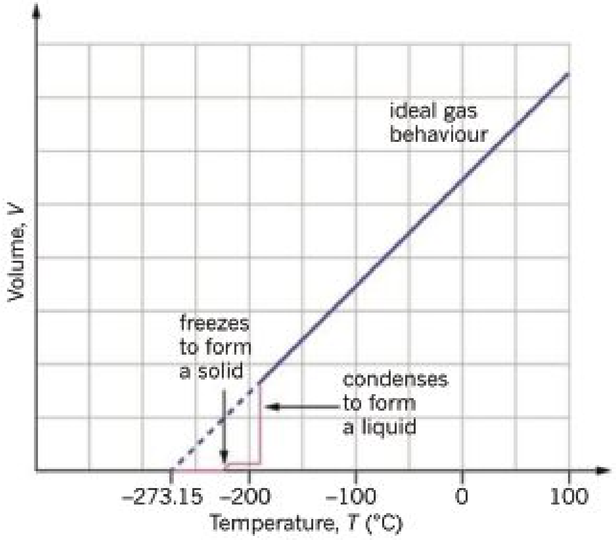


FIGURE 1.1.6 Gases have smaller volumes as they cool. This relationship is linear. Extrapolating (extending) the line to where the volume is zero gives a theoretical value of absolute zero.

PHYSICSFILE

Close to absolute zero

As temperatures get close to absolute zero, atoms start to behave in weird ways. Since the French physicist Guillaume Amontons first proposed the idea of an absolute lowest temperature in 1699, physicists have theorised about the effects of such a temperature and how it could be achieved. The laws of physics dictate that absolute zero itself can be approached but not reached. In 2003, researchers from NASA and MIT, in the United States of America, succeeded in cooling sodium atoms to one billionth of a degree above absolute zero. At this temperature, all elementary particles merge into a single state, losing their separate properties and behaving as a single 'super atom', a state first proposed by Einstein 70 years earlier.

THE LAWS OF THERMODYNAMICS

The topic of thermal physics involves the phenomena associated with the energy transfer between objects at different temperatures. Since the nineteenth century, scientists have developed four laws for this subject. The first two will be studied in this section.

The first, second and third laws had been known and understood for some time. Then another law was also determined. This final law was considered to be so important that it was decided to place it first, and so it is called the zeroth law of thermodynamics.

The zeroth law of thermodynamics

The zeroth law of thermodynamics relates to **thermal equilibrium** and **thermal contact** and allows temperature to be defined.

If two objects are in thermal contact, energy can flow between them. For example, if an ice cube is placed in a copper pan, the ice molecules are in thermal contact with the copper atoms. Assuming that the copper is warmer than the ice, thermal energy will flow from the copper to the ice.

Thermal equilibrium is when two objects in thermal contact stop having a flow of energy between them. If a frozen piece of steak is placed in a container of warm water, energy is transferred from the water to the steak. The steak gains energy and warms up. The water loses energy and cools down. Eventually the transfer of energy between the steak and water will stop. This point is called thermal equilibrium, and the steak and water will be at the same temperature.

> If objects A and B are each in thermal equilibrium with object C, then objects A and B are in thermal equilibrium with each other.

Two objects in thermal equilibrium with each other must be at the same temperature.

The first law of thermodynamics

The first law of thermodynamics states that energy simply changes from one form to another and the total internal energy in a system is constant. The internal energy of the system can be changed by heating or cooling, or by work being done on or by the system.

> Any change in the internal energy (ΔU) of a system is equal to the energy added by heating ($+Q$) or removed by cooling ($-Q$), minus the work done on ($-W$) or by ($+W$) the system.
>
> $\Delta U = Q - W$

The internal energy (U) of a system is defined as the total kinetic and potential energy of the system. As the average kinetic energy of a system is related to its temperature and the potential energy of the system is related to the state, then a change in the internal energy of a system means that either the temperature changes or the state changes.

If heat (Q) is added to the system, then the internal energy (U) rises by either increasing temperature or changing state from solid to liquid or liquid to gas. Similarly, if work (W) is done on a system, then the internal energy rises and the system will once again increase in temperature or will change state by melting or boiling. When heat is added to a system or work is done on a system, ΔU is positive.

If heat (Q) is removed from the system, then the internal energy (U) decreases by either decreasing temperature or changing state from liquid to solid or gas to liquid. Similarly, if work (W) is done by the system, then the internal energy decreases and the system will once again decrease in temperature or change state by condensing or solidifying. When heat is removed from a system or work is done by a system, ΔU is negative.

If heat is added to the system and work is done by the system, then whether the internal energy increases or decreases will depend on the magnitude of the energy into the system compared to the magnitude of the energy out of the system.

Worked example 1.1.1

CALCULATING THE CHANGE IN INTERNAL ENERGY

A 1 L beaker of water has 25 kJ of work done on it and also loses 30 kJ of thermal energy to the surroundings. What is the change in energy of the water?	
Thinking	**Working**
Heat is removed from the system, so Q is negative. Work is done on the system, so W is negative.	$\Delta U = Q - W$ $= -30 - (-25)$
Note that the units are kJ, so express the final answer in kJ.	$\Delta U = -5$ kJ

Worked example: Try yourself 1.1.1

CALCULATING THE CHANGE IN INTERNAL ENERGY

A student places a heating element and a paddle-wheel apparatus in an insulated container of water. She calculates that the heater transfers 2530 J of thermal energy to the water and the paddle does 240 J of work on the water. Calculate the change in internal energy of the water.

1.1 Review

SUMMARY

- The kinetic particle theory proposes that all matter is made of atoms or molecules (particles) that are in constant motion.
- In solids, the attractive and repulsive forces hold the particles in more or less fixed positions, usually in a regular arrangement or lattice. These particles are not completely still—they vibrate about average positions.
- In liquids, there is still a balance of attractive and repulsive forces between particles but the particles have more freedom to move around. Liquids maintain a fixed volume.
- In gases, the particle speeds are high enough that, when particles collide, the attractive forces are not strong enough to keep them close together. The repulsive forces cause the particles to move off in other directions.
- Internal energy refers to the total kinetic and potential energy of the particles within a substance.
- Temperature is related to the average kinetic energy of the particles in a substance.
- Heating is a process that always transfers thermal energy from a hotter substance to a colder substance.
- Temperatures can be measured in degrees Celsius (°C) or kelvin (K).
- Absolute zero is called simply 'zero kelvin' (0 K) and it is equal to −273.15°C.
- The size of each unit, 1°C or 1 K, is the same.
- To convert from Celsius to kelvin: add 273; to convert from kelvin to Celsius: subtract 273.
- The zeroth law of thermodynamics states that if objects A and B are each in thermal equilibrium with object C, then objects A and B are in thermal equilibrium with each other. A, B and C must be at the same temperature.
- The first law of thermodynamics states that energy simply changes from one form to another and the total energy in a system is constant.
- Any change in the internal energy (ΔU) of a system is equal to the energy added by heating ($+Q$) or removed by cooling ($-Q$), minus the work done on ($-W$) or by ($+W$) the system: $\Delta U = Q - W$.

1.1 Review *continued*

KEY QUESTIONS

1 Which of the following is true of a solid?
- **A** Particles are moving around freely.
- **B** Particles are not moving.
- **C** Particles are vibrating in constant motion.
- **D** A solid is not made up of particles.

2 **a** An uncooked chicken is placed into an oven that has been preheated to 180°C. Which of the following statements describe what happens as soon as the chicken is placed in the oven? (More than one answer is possible.)
- **A** Thermal energy flows from the chicken into the hot air.
- **B** The chicken and the air in the oven are in thermal equilibrium.
- **C** Thermal energy flows from the hot air into the chicken.
- **D** The chicken and the air in the oven are not in thermal equilibrium.

b A chicken is inside an oven that has been preheated to 180°C. The chicken has been cooking for one hour and its temperature is also 180°C. Which of the following statements best describes this scenario?
- **A** Thermal energy flows from the chicken into the hot air.
- **B** The chicken and the air in the oven are in thermal equilibrium.
- **C** Thermal energy flows from the hot air into the chicken.
- **D** The chicken and the air in the oven are not in thermal equilibrium

3 Which of the following temperature(s) cannot possibly exist? (More than one answer is possible.)
- **A** 1 000 000°C
- **B** −50°C
- **C** −50 K
- **D** −300°C

4 A tank of pure helium is cooled to its freezing point of −272.2°C. Describe the energy of the helium particles at this temperature.

5 Covert the following temperatures:
- **a** 30°C into kelvin
- **b** 375 K into degrees Celsius.

6 Tank A is filled with hydrogen gas at 0°C and another tank, B, is filled with hydrogen gas at 300 K. Describe the difference in the average kinetic energy of the hydrogen particles in each tank.

7 Sort the following temperatures from coldest to hottest:

freezing point of water
100 K
absolute zero
−180°C
10 K

8 A hot block of iron does 50 kJ of work on a cold floor. The block of iron also transfers 20 kJ of heat energy to the air. Calculate the change in energy (in kJ) of the iron block.

9 A chef vigorously stirs a pot of cold water and does 150 J of work on the water. The water also gains 75 J of thermal energy from the surroundings. Calculate the change in energy of the water.

10 A scientist very carefully does mechanical work on a container of liquid sodium. The liquid sodium loses 300 J of energy to its surroundings but gains 250 J of energy overall. How much work did the scientist do?

1.2 Specific heat capacity

A small amount of water in a kettle will experience a greater change in temperature than a larger volume if heated for the same time. A metal object left in the sunshine get hotter faster than a wooden object. Large heaters warm rooms faster than small ones.

These simple observations suggest that the mass, material and the amount of energy transferred influence any change of temperature.

CHANGING TEMPERATURE

The temperature of a substance is a measure of the average kinetic energy of the particles inside the substance. To increase the temperature of the substance, the kinetic energy of its particles must increase. This happens when heat is transferred to that substance. The amount the temperature increases depends on a number of factors.

The greater the mass of a substance, the greater the energy required to change the kinetic energy of all the particles. So, the heat required to raise the temperature by a given amount is proportional to the mass of the substance.

$\Delta Q \propto m$

where ΔQ is the heat energy transferred in joules (J)

m is the mass of material being heated in kilograms (kg).

The more heat that is transferred to a substance, the more the temperature of that substance increases. The amount of energy transferred is therefore proportional to the change in temperature.

$\Delta Q \propto \Delta T$

where ΔT is the change in temperature in °C or K.

Heating experiments using different materials will confirm that these relationships hold true regardless of the material being heated. However, heating the same masses of different materials will show that the amount of energy required to heat a given mass of a material through a particular temperature change also depends on the nature of the material being heated. For example, a volume of water requires more energy to change its temperature by a given amount compared with the same volume of methylated spirits. For some materials, temperature change occurs more easily than for others.

The **specific heat capacity** of a material, c, is the amount of energy that must be transferred to change the temperature of 1 kg of the material by 1°C or 1 K.

Combining these observations, the amount of energy added to or removed from the substance is proportional to the change in its temperature, its mass and its specific heat capacity (provided a material does not change state). The **specific heat capacity** of a material changes when the material changes state.

As an equation:

$Q = mc\Delta T$

where Q is the heat energy transferred in joules (J)

m is the mass in kilograms (kg)

ΔT is the change in temperature in °C or K

c is the specific heat capacity of the material (J kg^{-1} K^{-1}).

Table 1.2.1 lists the specific heat capacities for some common materials. You can see that it also lists the average value for the human body, which takes into account the various materials within the body and the percentage that each material contributes to the body's total mass.

Material	c (J kg^{-1} K^{-1})
human body	3500
methylated spirits	2500
air	1000
aluminium	900
glass	840
iron	440
copper	390
brass	370
lead	130
mercury	140
ice (water)	2100
liquid water	4200
steam (water)	2000

TABLE 1.2.1 Approximate specific heat capacities of common substances.

Worked example 1.2.1

CALCULATIONS USING SPECIFIC HEAT CAPACITY

A hot water tank contains 135 L of water. Initially the water is at 20°C. Calculate the amount of energy that must be transferred to the water to raise the temperature to 70°C.	
Thinking	**Working**
Calculate the mass of water. 1 L of water = 1 kg	Volume = 135 L so mass of water = 135 kg
ΔT = final temperature − initial temperature	$\Delta T = 70 - 20 = 50$°C
From the table of specific heat capacities on page 10, $c_{\text{water}} = 4200\ \text{J kg}^{-1}\ \text{K}^{-1}$. Use the equation $Q = mc\Delta T$.	$Q = mc\Delta T$ $= 135 \times 4200 \times 50$ $= 28\,350\,000$ J $= 28$ MJ

Worked example: Try yourself 1.2.1

CALCULATIONS USING SPECIFIC HEAT CAPACITY

A bath contains 75 L of water. Initially the water is at 50°C. Calculate the amount of energy that must be transferred from the water to cool the bath to 30°C.

Worked example 1.2.2

COMPARING SPECIFIC HEAT CAPACITIES

Different states of matter of the same substance have different specific heat capacities. What is the ratio of the specific heat capacity of liquid water to that of ice?	
Thinking	**Working**
See Table 1.2.1 for the specific heat capacities of water in different states.	$c_{\text{water}} = 4200\ \text{J kg}^{-1}\ \text{K}^{-1}$ $c_{\text{ice}} = 2100\ \text{J kg}^{-1}\ \text{K}^{-1}$
Divide the specific heat of water by the specific heat of ice.	Ratio $= \frac{c_{\text{water}}}{c_{\text{ice}}}$ $= \frac{4200}{2100}$
Note that ratios have no units since the unit of each quantity is the same and cancels out.	Ratio = 2

Worked example: Try yourself 1.2.2

COMPARING SPECIFIC HEAT CAPACITIES

What is the ratio of the specific heat capacity of liquid water to that of steam?

PHYSICSFILE

The mass of water

Since water is a familiar material, many of the examples in this section use it as the liquid being heated. One kilogram of pure water has a volume of 1 litre at 4°C.

PHYSICSFILE

Specific heat capacity of water

One of the notable values in the table of specific heat capacities is the high value for water. It is 10 times, or an order of magnitude, higher than those of most metals listed. The specific heat capacity of water is higher than those of most common materials. As a result, water makes a very useful cooling and heat storage agent, and is used in areas such as generator cooling towers and car-engine radiators.

Life on Earth also depends on the specific heat capacity of water. About 70% of the Earth's surface is covered by water, and these water bodies can absorb large quantities of thermal energy without great changes in temperature. Oceans both heat up and cool down more slowly than the land areas next to them. This helps to maintain a relatively stable range of temperatures for life on Earth.

Scientists are now monitoring the temperatures of the deep oceans in order to determine how the ability of oceans to store large amounts of energy may affect climate change.

1.2 Review

SUMMARY

- When heat is transferred to or from a system or object, the temperature change depends upon the amount of energy transferred, the mass of the material(s) and the specific heat capacity of the material(s):
  $Q = mc\Delta T$
  where Q is the heat energy transferred in joules (J)
  m is the mass of material being heated in kilograms (kg)
  ΔT is the change in temperature (°C or K)
  c is the specific heat capacity of the material (J $\text{kg}^{-1}\ \text{K}^{-1}$).
- A substance will have different specific heat capacities at different states (solid, liquid, gas).

KEY QUESTIONS

1 Equal masses of water and aluminium are heated through the same temperature range. Using the values of c from Table 1.2.1 on page 10, which material requires the most energy to achieve this result?

2 Which has the most thermal energy: 10 kg of iron at 20°C or 10 kg of aluminium at 20°C?

3 100 mL of water is heated to change its temperature from 15°C to 20°C. How much energy is transferred to the water to achieve this temperature change?

4 150 mL of water is heated from 10°C to 50°C. What amount of energy is required for this temperature change to occur?

5 For a 1 kg block of aluminium, x J of energy are needed to raise the temperature by 10°C. How much energy, in J, is needed to raise the temperature by 20°C?

6 Equal amounts of energy are absorbed by equal masses of aluminium and water. What is the ratio of the temperature rise of the aluminium to that of water?

7 Which one or more of the following statements about specific heat capacity is true?
  A All materials have the same specific heat capacity when in solid form.
  B The specific heat capacity of a liquid form of a material is different from that of the solid and gas forms.
  C Good conductors of heat generally have high specific heat capacities.
  D Specific heat capacity is independent of temperature.

8 If 4.0 kJ of energy is required to raise the temperature of 1.0 kg of paraffin by 2.0°C, how much energy (in kJ) is required to raise the temperature of 5.0 kg of paraffin by 1.0°C?

9 A cup holds 250 mL of water at 20°C. 10.5 kJ of heat energy is transferred to the water. What temperature does the water reach after the heat is transferred?

10 A block of iron is left to cool. After cooling for a short time, 13.2 kJ of energy has been transferred away from the block of iron and its temperature has decreased by 30°C. What is the mass of the block of iron?

1.3 Latent heat

If water is heated, its temperature will rise. If enough energy is transferred to the water, eventually the water will boil. The water changes state (from liquid to gas). The **latent heat** is the energy released or absorbed during a change of state. Latent means hidden or unseen. While a substance changes state, its temperature remains constant. The energy used in, say, melting ice into water is hidden in the sense that the temperature doesn't rise while the change of state is occurring.

ENERGY AND CHANGE OF STATE

Look at the heating curve for water shown in Figure 1.3.1. This graph shows how the temperature of water changes as energy is added at a constant rate. Although the rate at which the energy is added is constant, the increase in temperature is not always constant. There are sections of increasing temperature, and sections where the temperature remains unchanged (the horizontal sections) while the material changes state. Temperature remains constant during the change in state from ice to liquid water and again from liquid water to steam.

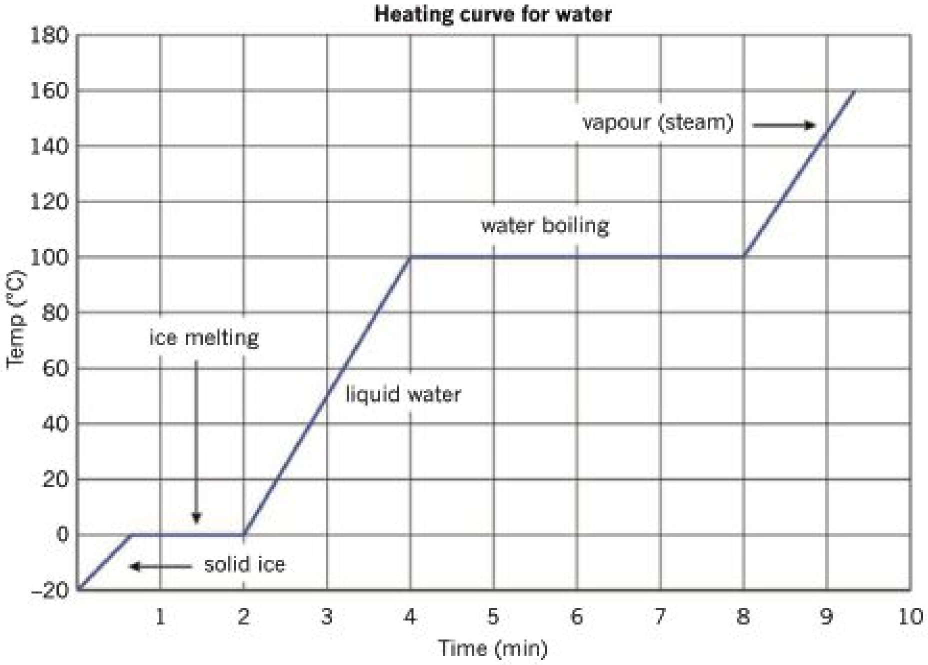


FIGURE 1.3.1 A heating curve for water.

Latent heat

The energy needed to change the state of a substance (e.g. solid to liquid, liquid to gas) is called latent heat. Latent heat is the 'hidden' energy that has to be added or removed from a material in order for the material to change state.

The latent heat is calculated using the equation:

$Q = mL$

where Q is the heat energy transferred in joules (J)

m is the mass in kilograms (kg)

L is the latent heat (J kg^{-1}).

LATENT HEAT OF FUSION (MELTING)

As energy is transferred to a solid, the temperature of the solid increases. The particles within the solid gain internal energy (as kinetic energy and some potential energy) and their speed of vibration increases. At the point where the solid begins to melt, the particles move further apart, reducing the strength of the bonds holding them in place. At this point, instead of increasing the temperature, the extra energy increases the potential energy of the particles, reducing the interparticle or intermolecular forces. No change in temperature occurs as all the extra energy supplied is used in reducing these forces between particles.

The amount of energy required to melt a solid is exactly the same as the amount of potential energy released when the liquid re-forms into a solid. It is termed the **latent heat of fusion**.

The amount of energy required will depend on the particular solid.

For a given mass of a substance:

heat energy transferred = mass of substance × specific latent heat of fusion

$Q = mL_{\text{fusion}}$

where Q is the heat energy transferred in joules (J)

m is the mass in kilograms (kg)

L_{fusion} is the latent heat of fusion in J kg^{-1}.

It takes almost 80 times as much energy to turn 1 kg of ice into water (with no temperature change) as it does to raise the temperature of 1 kg of water by 1°C. It takes a lot more energy to overcome the large intermolecular forces within the ice than it does to simply add kinetic energy in raising the temperature.

The latent heats of fusion for some common materials are shown in Table 1.3.1.

Substance	Melting point (°C)	L_{fusion} (J kg^{-1})
water	0	3.34×10^5
oxygen	−219	0.14×10^5
lead	327	0.25×10^5
ethanol	−114	1.05×10^5
silver	961	0.88×10^5

TABLE 1.3.1 The latent heats of fusion for some common materials.

Worked example 1.3.1

LATENT HEAT OF FUSION

<table>
<tr><td colspan="2">How much energy must be removed from 2.5 L of water at 0°C to produce a block of ice at 0°C? Express your answer in kJ.</td></tr>
<tr><td>Thinking</td><td>Working</td></tr>
<tr><td>Cooling from liquid to solid involves the latent heat of fusion, where the energy is removed from the water.
Calculate the mass of water involved.</td><td>1 L of water = 1 kg, so 2.5 L = 2.5 kg</td></tr>
<tr><td>Use Table 1.3.1 to find the latent heat of fusion for water.</td><td>$L_{\text{fusion}} = 3.34 \times 10^5$ J kg^{-1}</td></tr>
<tr><td>Use the equation $Q = mL_{\text{fusion}}$.</td><td>$Q = mL_{\text{fusion}}$
$= 2.5 \times 3.34 \times 10^5$
$= 8.35 \times 10^5$ J</td></tr>
<tr><td>Convert to kJ.</td><td>$Q = 8.35 \times 10^2$ kJ</td></tr>
</table>

Worked example: Try yourself 1.3.1

LATENT HEAT OF FUSION

How much energy must be removed from 5.5 kg of liquid lead at 327°C to produce a block of solid lead at 327°C? Express your answer in kJ.

LATENT HEAT OF VAPORISATION (BOILING)

It takes much more energy to convert a liquid to a gas than it does to convert a solid to a liquid. This is because, to convert to a gas, the intermolecular bonds must be broken. During the change of state, the energy supplied is used solely in overcoming intermolecular bonds. The temperature will not rise until all of the material in the liquid state is converted to a gas, assuming that the liquid is evenly heated. For example, when liquid water is heated to boiling point, a large amount of energy is required to change its state from liquid to steam (gas). The temperature will remain at 100°C until all of the water has turned into steam. Once the water is completely converted to steam, then the temperature can start to rise again.

The amount of energy required to change a liquid to a gas is exactly the same as the potential energy released when the gas returns to a liquid. It is called the **latent heat of vaporisation**.

The amount of energy required will depend on the particular substance.

For a given mass of a substance:

heat energy transferred = mass of substance × specific latent heat of vaporisation

$Q = mL_{\text{vapour}}$

where Q is the heat energy transferred in joules (J)

m is the mass in kilograms (kg)

L_{vapour} is the latent heat of vaporisation (J kg^{-1}).

Note that, in just about every case, the latent heat of vaporisation of a substance will be different to the latent heat of fusion for that substance. Some latent heat of vaporisation values are listed in Table 1.3.2.

In many instances, it is necessary to consider the energy required to heat a substance and also change its state. Problems like this are solved by considering the rise in temperature separately from the change of state.

Substance	Melting point (°C)	L_{vapour} (J kg^{-1})
water	100	22.5×10^5
oxygen	−183	2.2×10^5
lead	1750	9.0×10^5
ethanol	78	8.7×10^5
silver	2193	23.0×10^5

TABLE 1.3.2 The latent heat of vaporisation of some common materials.

Worked example 1.3.2

CHANGE IN TEMPERATURE AND STATE

50 mL of water is heated from a room temperature of 20°C to its boiling point at 100°C. It is boiled at this temperature until it is completely evaporated. How much energy in total was required to raise the temperature and boil the water?

Thinking	**Working**
Calculate the mass of water involved.	50 mL of water = 0.05 kg
Find the specific heat capacity of water (see Table 1.2.1).	$c = 4200$ J kg^{-1} K^{-1}
Use the equation $Q = mc\Delta T$ to calculate the heat energy required to change the temperature of water from 20°C to 100°C.	$Q = mc\Delta T$ $= 0.05 \times 4200 \times (100 - 20)$ $= 16800$ J
Find the specific latent heat of vaporisation of water.	$L_{\text{vapour}} = 22.5 \times 10^5$ J kg^{-1}
Use the equation $Q = mL_{\text{vapour}}$ to calculate the latent heat required to boil water.	$Q = mL_{\text{vapour}}$ $= 0.05 \times 22.5 \times 10^5$ $= 112500$ J
Find the total energy required to raise the temperature and change the state of the water.	Total $Q = 16800 + 112500$ $= 129300$ J (or 1.29×10^5 J)

PHYSICSFILE

Extinguishing fire

The latent heat of vaporisation of water is very high. This is due to the molecular structure of the water. This characteristic of water makes it very useful for extinguishing fires. That's because water can absorb vast amounts of thermal energy before it evaporates. By pouring water onto a fire, energy is transferred away from the fire to heat the water. Then, even more (in fact much more) heat is transferred away from the fire to convert the water into steam.

Worked example: Try yourself 1.3.2

CHANGE IN TEMPERATURE AND STATE

3 L of water is heated from a fridge temperature of 4°C to its boiling point at 100°C. It is boiled at this temperature until it is completely evaporated. How much energy in total was required to raise the temperature and boil the water?

EVAPORATION AND COOLING

If you spill some water on the floor then come back in a couple of hours, the water will probably be gone. It will have evaporated. It has changed from a liquid into a vapour at room temperature in a process called **evaporation**. The reason for this is that the water particles, if they have sufficient energy, are able to escape through the surface of the liquid into the air. Over time, no liquid remains.

Evaporation is more noticeable in **volatile** liquids such as methylated spirits, mineral turpentine, perfume and liquid paper. The surface bonds are weaker in these liquids and they evaporate rapidly. This is why you should never leave the lid off bottles of these liquids. They are often stored in narrow-necked bottles for this reason.

The rate of evaporation of a liquid depends on:

- the volatility of the liquid: more-volatile liquids evaporate faster
- the surface area: greater evaporation occurs when greater surface areas are exposed to the air
- the temperature: hotter liquids evaporate faster
- the humidity: less evaporation occurs in more humid conditions
- air movement: if a breeze is blowing over the liquid's surface, evaporation is more rapid.

Whenever evaporation occurs, higher-energy particles escape the surface of the liquid, leaving the lower-energy particles behind, as is shown in Figure 1.3.2. As a result, the average kinetic energy of the particles remaining in the liquid drops and the temperature decreases. Humans use this cooling principle when sweating to stay cool. When rubbing alcohol is dabbed on your arm before an injection, the cooling of the volatile liquid numbs your skin.

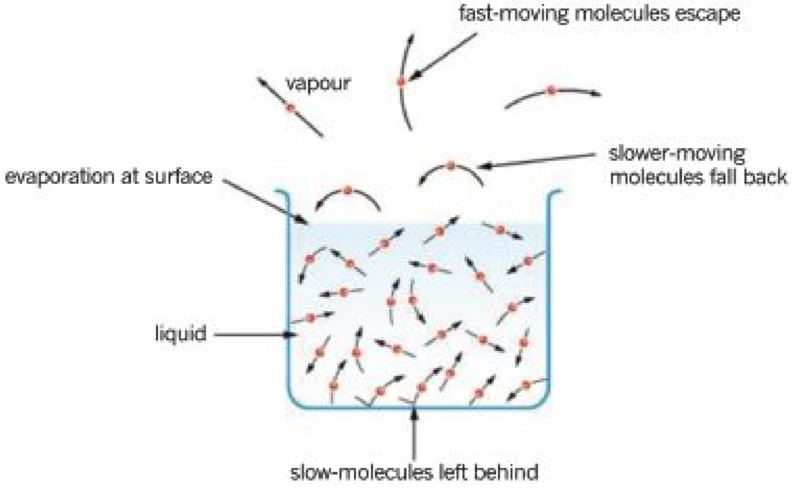


FIGURE 1.3.2 Fast-moving molecules with high kinetic energy can escape the liquid, leaving molecules with lower kinetic energy behind.

Other applications of evaporation such as evaporative coolers will be discussed in more detail in Chapter 2.

1.3 Review

SUMMARY

- When a solid material changes state, energy is needed to separate the particles by overcoming the attractive forces between the particles.
- Latent heat is the energy required to change the state of 1 kg of material at a constant temperature.
- In general, for any mass of material the energy required (or released) is
  $Q = mL$
  where Q is the energy transferred in joules (J)
  m is the mass in kilograms (kg)
  L is the latent heat (J kg^{-1}).
- The latent heat of fusion, L_{fusion}, is the energy required to change 1 kg of a material between the solid and liquid states.
- The latent heat of vaporisation, L_{vapour}, is the energy required to change 1 kg of a material between the liquid and gaseous states.
- The latent heat of fusion of a material will be different to (and usually less than) the latent heat of vaporisation for that material.
- Evaporation is when a liquid turns into gas at room temperature. The temperature of the liquid falls as this occurs.
- The rate of evaporation depends on the volatility, temperature and surface area of the liquid and the presence of a breeze.

KEY QUESTIONS

Refer to the values in Table 1.3.1 and Table 1.3.2 on pages 14 and 15. You may also need to refer to Table 1.2.1 on page 10.

The following information relates to questions 1 to 5.

Figure 1.3.3 represents the heating curve for mercury, a metal that is a liquid at normal room temperature. Thermal energy is added to 10 g of solid mercury, initially at a temperature of −39°C, until all of the mercury has evaporated.

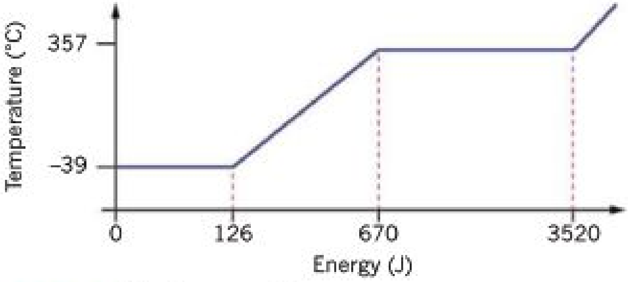


FIGURE 1.3.3 A heating curve of mercury.

1. Why does the temperature remain constant during the first part of the graph?
2. What is the melting point of mercury, in degrees Celsius?
3. What is the boiling point of mercury, in degrees Celsius?
4. From the graph, what is the latent heat of fusion of mercury?
5. From the graph, what is the latent heat of vaporisation of mercury?
6. How much heat energy must be transferred away from 100 g of steam at 100°C to change it completely to a liquid?
7. How many kJ of energy are required to melt exactly 100 g of ice initially at −4.00°C? Assume no loss of energy to surroundings.
8. Which of the following explains why hot water in a spa-pool evaporates more rapidly than cold water?
   A Hot water molecules have less energy than cold water molecules.
   B Hot water molecules have more energy than cold water molecules.
   C Hot water forms water vapour and bubbles to the surface.
   D Hot water dissolves into the pool material more rapidly.
9. A painter spills some mineral turpentine onto a concrete floor. After a minute, most of the liquid is gone. Which of the following is correct?
   A Most of the liquid has boiled, becoming hotter.
   B Most of the liquid has evaporated and the remaining liquid becomes warmer as it does so.
   C Most of the liquid has evaporated with no change in temperature of the remaining liquid.
   D Most of the liquid has evaporated and the remaining liquid becomes colder as it does so.

1.4 Conduction

FIGURE 1.4.1 Emperor penguin chicks avoid heat loss through conduction by sitting on the adult's feet. In this way they avoid contact with the ice.

If two objects are at different temperatures and are in thermal contact (that is, they can exchange energy via heat processes), then thermal energy will transfer from the hotter object to the cooler object. Figure 1.4.1 shows how, by preventing the chick's thermal contact with the cold ice, this adult is able to protect the vulnerable penguin offspring.

There are three possible means by which heat can be transferred:

- conduction
- convection
- radiation.

This section focuses on conduction.

CONDUCTORS AND INSULATORS

Conduction is the process by which heat is transferred from one place to another without the net movement of particles (atoms or molecules). Conduction can occur within a material or between materials that are in thermal contact. For example, if one end of a steel rod is placed in a fire, heat will travel along the rod so that the far end of the rod will also heat up; or if a person holds an ice cube, then heat will travel from their hand to the ice.

While all materials will conduct heat to some extent, this process is most significant in solids. It is important in liquids but plays a lesser role in the movement of energy in gases.

Materials that conduct heat readily are referred to as good **conductors**. Materials that are poor conductors of heat are referred to as **insulators**. An example of a good conductor and a good insulator can be seen in Figure 1.4.2.

In secondary physics, the terms 'conductor' and 'insulator' are used in the context of both electricity and heating processes. What makes a material a good conductor of heat doesn't necessarily make it a good conductor of electricity. The two types of conduction are related but it's important not to confuse the two processes. A material's ability to conduct heat depends on how conduction occurs within the material.

Conduction can happen in two ways:

- energy transfer through molecular or atomic collisions
- energy transfer by free electrons.

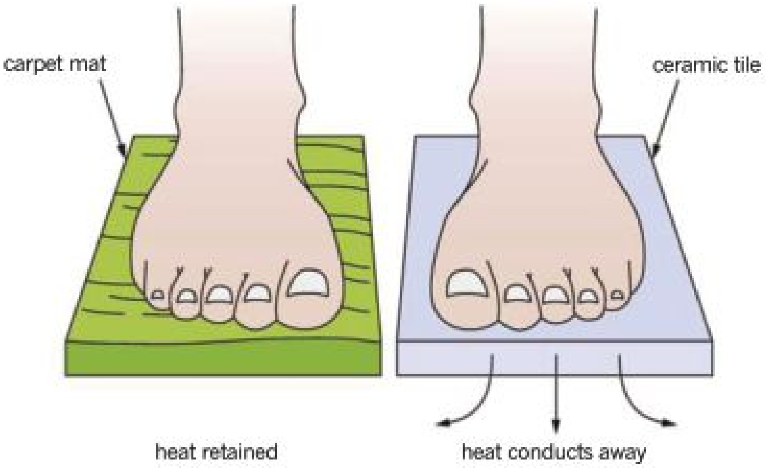


FIGURE 1.4.2 Ceramic floor tiles are good conductors of heat. They conduct heat away from the foot readily and so your feet feel cold on tiles. The carpet mat is a thermal insulator. Thermal energy from the foot is not transferred away as quickly and so your foot doesn't feel as cold.

THERMAL TRANSFER BY COLLISION

The kinetic particle model explains that particles in a solid substance are constantly vibrating within the material structure and so interact with neighbouring particles. If one part of the material is heated, then the particles in that region will vibrate more rapidly. Interactions with neighbouring particles will pass on this kinetic energy throughout the system via the bonds between the particles (see Figure 1.4.3).

The process can be quite slow since the mass of the particles is relatively large and the vibrational velocities are fairly low. Materials for which this method of conduction is the only means of heat transfer are likely to be poor conductors of heat or even thermal insulators. Materials such as glass, wood and paper are poor conductors of heat.

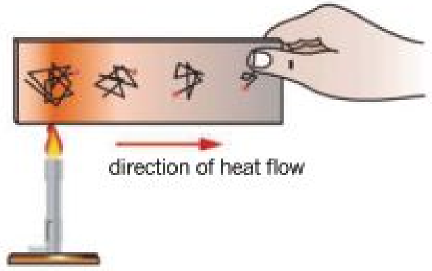


FIGURE 1.4.3 Thermal energy is passed on by collisions between adjacent particles.

THERMAL TRANSFER BY FREE ELECTRONS

Some materials, particularly metals, have electrons that are not directly involved in any one particular chemical bond. Therefore, these electrons are free to move throughout the lattice of positive ions.

For example, if a metal is heated, then not only will the positive ions within the metal gain extra energy but so will these free electrons. As the electron's mass is considerably less than the positive ions, even a small energy gain will result in a very large gain in velocity. Consequently, these free electrons provide a means by which heat can be quickly transferred throughout the whole of the material. It is therefore no surprise that metals, which are good electrical conductors because of these free electrons, are also good thermal conductors.

THERMAL CONDUCTIVITY

Thermal conductivity describes the ability of a material to conduct heat. It is temperature dependent and is measured in watts per metre kelvin (W m^{-1} K^{-1}). Table 1.4.1 highlights the difference in conductivity in metals compared with other substances.

Material	Conductivity (W m^{-1} K^{-1})
silver	420
copper	380
aluminium	240
steel	60
ice	2.2
brick, glass	≈ 1
concrete	≈ 1 (depending on composition)
water	0.6
human tissue	0.2
wood	0.15
polystyrene	0.08
paper	0.06
fibreglass	0.04
air	0.025

TABLE 1.4.1 Thermal conductivities of some common materials.

Factors affecting thermal conduction

The rate at which heat will be transferred through a system depends on the:

- nature of the material. The larger a material's thermal conductivity, the more rapidly it will conduct heat energy.
- temperature difference between the two objects. A greater temperature difference will result in a faster rate of energy transfer.
- thickness of the material. Thicker materials require a greater number of collisions between particles or movement of electrons to transfer energy from one side to the other.
- surface area. Increasing the surface area relative to the volume of a system increases the number of particles involved in the transfer process, increasing the rate of conduction.

The rate at which heat is transferred is measured in joules per second (J s^{-1}), or watts (W).

PHYSICSFILE

Igloos

It seems strange that an igloo can keep a person warm when ice is so cold. Igloos are constructed from compressed snow that contains many air pockets. The air in these pockets is a poor conductor of heat, which means heat inside the igloo is not easily transferred away. The body heat of the occupant, as well as that of his or her small heat source, is trapped inside the igloo and able to keep them warm.

EXTENSION

Thermal conductivity

The rate of energy transfer by conduction (energy per unit time) through a material can be calculated using:

$$\frac{Q}{t} = \frac{kA\Delta T}{L}$$

where $\frac{Q}{t}$ is the heat energy transferred in joules (J) per unit time, t, in seconds (s)

k is the thermal conductivity of the material (W m^{-1} K^{-1})

A is the surface area in metres squared perpendicular to the direction of heat flow (m^2)

ΔT is the temperature difference across the material in kelvin or degrees Celsius (K or °C)

L is the thickness of the material through which the heat is being transferred in metres (m).

Designers use this relationship when calculating the insulating ability of the fill inside clothing such as parkas. Architects and builders use it to calculate the efficiency of building insulation.

Guidelines exist to ensure the efficiency of insulating materials. Building materials that limit the transfer of heat help to keep houses warm in winter and cool in summer. This saves money and helps to reduce carbon dioxide emissions from the use of gas or electricity to heat houses.

1.4 Review

SUMMARY

- Conduction is the process of heat transfer within a material or between materials without the overall transfer of the substance itself.
- All materials will conduct heat to a greater or lesser degree. Materials that readily conduct heat are called good thermal conductors. Materials that conduct heat poorly are called thermal insulators.
- Whether a material is a good conductor depends on the method of conduction:
  - Heat transfer by molecular collisions alone occurs in poor to very poor conductors.
  - Heat transfer by molecular collisions and free electrons occurs in good to very good conductors.
- The rate of conduction depends on the temperature difference between two materials, the thickness of the material, the surface area and the nature of the material.

KEY QUESTIONS

1 Explain why the process of conduction by molecular collision is slow.

2 Why are metals more likely to conduct heat than wood?

3 List the properties of a material that affect its ability to conduct heat.

4 Stainless steel saucepans are often manufactured with a copper base. What is the most likely reason for this?

5 One way of making a house energy-efficient is to use double-glazed windows. These consist of two panes of glass with air trapped in between the panes. On a hot day, the energy from the hot air outside the house is not able to penetrate the air gap and so the house stays cool.

Which of the following best explains why double-glazing works?

A Air is a conductor of heat and so the thermal energy is able to pass through.

B Air is a conductor of heat and so the thermal energy is not able to pass through.

C Air is an insulator of heat and so the thermal energy is not able to pass through.

D Air is an insulator of heat and so the thermal energy is able to pass through.

6 How does a down-filled quilt keep a person warm in winter?

7 Fibreglass insulation batts are thick and lightweight, and they make a house more energy-efficient. On a cold July night, the external temperature in the roof of an insulated house is 6°C. The air temperature near the ceiling inside the house is 20°C.

Complete the paragraph below by choosing the correct response from the choices given in brackets to explain the effects of ceiling insulation in this situation.

The insulation batts stop the thermal energy from [**escaping/entering**] the house. The air in the batts has [**high/low**] conductivity and the thermal energy is [**able/not able**] to from escape from the house.

8 On a cold day, the plastic or rubber handles of a bicycle feel much warmer than the metal surfaces. Explain this in terms of the thermal conductivity of each material.

1.5 Convection

This section will investigate convection, which is the transfer of thermal energy within a fluid (liquid or a gas) by the movement of hot areas from one place to another. Unlike other forms of heat transfer such as conduction and radiation, convection involves the mass movement of particles within a system over a distance that can be quite considerable.

HEATING BY CONVECTION

Although liquids and gases are generally not very good conductors of thermal energy, heat can be transferred quite quickly through liquids and gases by convection. Unlike other forms of thermal energy transfer, **convection** involves the mass movement of particles within a system over a distance.

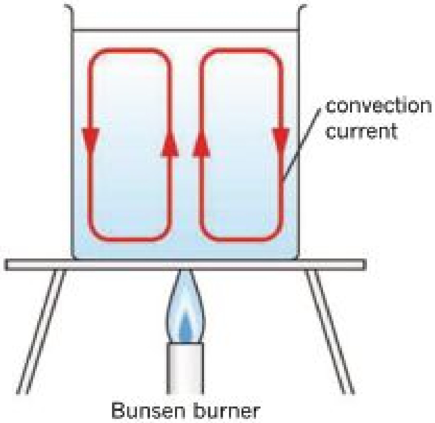


FIGURE 1.5.1 When a liquid or gas is heated, it becomes hotter and less dense so will rise. The colder, denser fluid will fall. As this fluid heats up, it in turn will rise, creating a convection current.

As a fluid is heated, the particles within it gain kinetic energy and push apart due to the increased vibration of the particles. This causes the density of the heated fluid to decrease and the heated fluid rises. Colder fluid, with slower moving particles, is more dense and heavier and hence falls, moving in to take the place of the warmer fluid. A convection current forms when there is warm fluid rising and cool fluid falling. This action can be seen in Figure 1.5.1. Upwellings in oceans, wind and weather patterns are at least partially due to convection on a very large scale.

It is difficult to quantify the thermal energy transferred via convection but some estimates can be made. The rate at which convection will occur is affected by:

- the temperature difference between the heat source and the convective fluid
- the surface area exposed to the convective fluid.

In a container, the effectiveness of convection to transfer heat depends on the placement of the source of heat. For example, the heating element in a kettle is always found near the bottom of the kettle. From this position, convection currents form throughout the water to heat it more effectively (see Figure 1.5.2(a)). If the heating element is placed near the top of the kettle, convection currents form only near the top. This is because the hotter water is less dense than the cooler water below and will remain near the top. Convection currents will not form throughout the water (see Figure 1.5.2(b)).

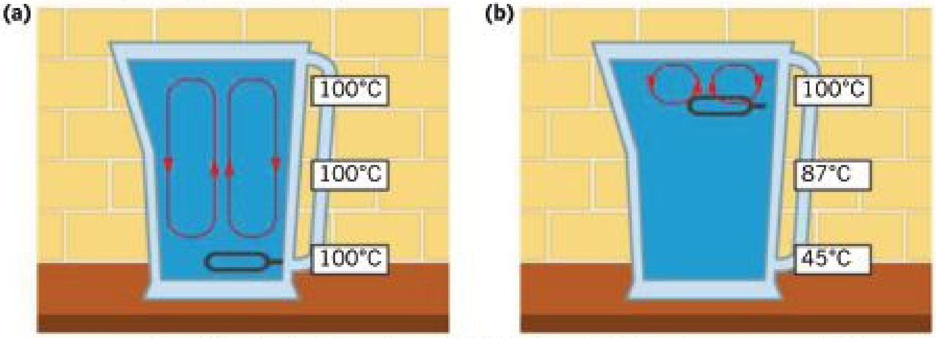


FIGURE 1.5.2 (a) By placing the heating element at the bottom of a kettle, the water near the bottom is heated and rises, forming convection currents throughout the entire depth of the water. (b) If the heating element is placed near the top of the kettle, the convection currents form near the top and heat transfer is slower.

There are two main causes of convection:

- forced convection, for example ducted heating in which air is heated and then blown into a room
- natural convection, such as that illustrated in Figure 1.5.1, when a fluid rises as it is heated.

FIGURE 1.5.3 The thunderheads of summer storms are a very visible indication of natural convection in action.

A dramatic example of natural convection is the thunderhead clouds of summer storms (as seen in Figure 1.5.3), which form when hot, humid air from natural convection currents is carried rapidly upwards into the cooler upper atmosphere.

PHYSICS IN ACTION

Wind chill

Convective effects are the main means of heat transfer that lead to the 'wind chill' factor. The wind blows away the thin layer of relatively still air near the skin that would normally act as a partial insulator in still air. Cooler air comes in closer contact with the skin and heat loss increases. It feels as if the 'effective' temperature of the surrounding air has decreased. Skiers can experience similar effects simply from the wind created by their own motion.

In cold climates the wind chill factor can become an important factor to consider. The chilling effect is even more dramatic when the body or clothing is wet, increasing evaporative cooling. Bushwalkers look for clothing that dries rapidly after rain and which carries moisture from the perspiration of heavy exertion away from the skin.

PHYSICSFILE

Paragliders

Paragliders fly by sitting in a harness suspended beneath a fabric wing. They gain altitude by catching thermals. Thermals are columns of rising hot air created by dark regions on the ground that have been heated up by the Sun. Roads, rock faces and ploughed fields are good at creating thermals.

In 2007, Polish paraglider Ewa Wisnierska was practising in NSW for a competition when she was caught in an intense thermal updraught during a storm. She reached an altitude of almost 10 km. Fortunately, she lost altitude, and landed about 60 km from where she started, where her crew found and rescued her. Ewa is now a paragliding instructor in Germany.

FIGURE 1.5.4 Paragliders can gain altitude by finding a thermal. These are areas of rising hot air created by hot regions on the ground. These paragliders are flying near Bright, Victoria.

1.5 Review

SUMMARY

- Convection is the transfer of heat within a fluid (liquid or gas).
- Convection involves the mass movement of particles within a system over a distance.
- A convection current forms when there is warm fluid rising and cool fluid falling.

KEY QUESTIONS

1. Through what states of matter can convection occur?
2. In which direction does the transfer of heat in a convection current initially flow?
3. Pilots of glider aircraft or hang-gliders, some birds such as eagles and some insects rely on 'thermals' to give them extra lift. Explain how these rising columns of air are established.
4. Which of the following statements about liquids and gases is true?

   A Liquids and gases can transfer heat quite slowly through convection even though liquids and gases are very good conductors of thermal energy.

   B Liquids and gases can transfer heat quite quickly through conduction because liquids and gases are very good conductors of thermal energy.

   C Liquids and gases can transfer heat quite quickly through convection because liquids and gases are very good conductors of thermal energy.

   D Liquids and gases can transfer heat quite quickly through convection even though liquids and gases are not very good conductors of thermal energy.
5. Convection is referred to as a method of heat transfer through fluids. Explain whether it is possible for solids to pass on their heat energy by convection.
6. On a hot day, the top layer of water in a swimming pool can heat up while the lower, deeper parts of the water can remain quite cold. Explain, using the concept of convection, why this happens.

1.6 Radiation

Both convection and conduction involve the transfer of heat through matter. Life on Earth depends upon the transfer of energy from the Sun through the near-vacuum of space. If heat could only be transferred by the action of particles, then the Sun's energy would never reach Earth. Radiation is a means of transfer of heat without the movement of matter.

RADIATION

In this context, **radiation** is a shortened form of electromagnetic radiation, which includes visible, ultraviolet and infrared light. Together with other forms of light, these make up the **electromagnetic spectrum**.

The transfer of heat from one place to another without the movement of particles is by electromagnetic radiation. Electromagnetic radiation travels at the speed of light. When electromagnetic radiation (light) hits an object, it will be partially reflected, partially transmitted and partially **absorbed**. The absorbed part transfers thermal energy to the absorbing object and causes a rise in temperature. When you hold a marshmallow by an open fire, you are using radiation to toast the marshmallow, as shown in Figure 1.6.1.

FIGURE 1.6.1 Heat transfer from the flame to the marshmallow is an example of radiation.

Electromagnetic radiation is **emitted** by all objects that are at a temperature above absolute zero (0 K or −273°C). The **wavelength** and **frequency** of the emitted radiation depends on the internal energy of the object. The higher the temperature of the object, the higher the frequency and the shorter the wavelength of the radiation emitted. This can be seen in Figure 1.6.2.

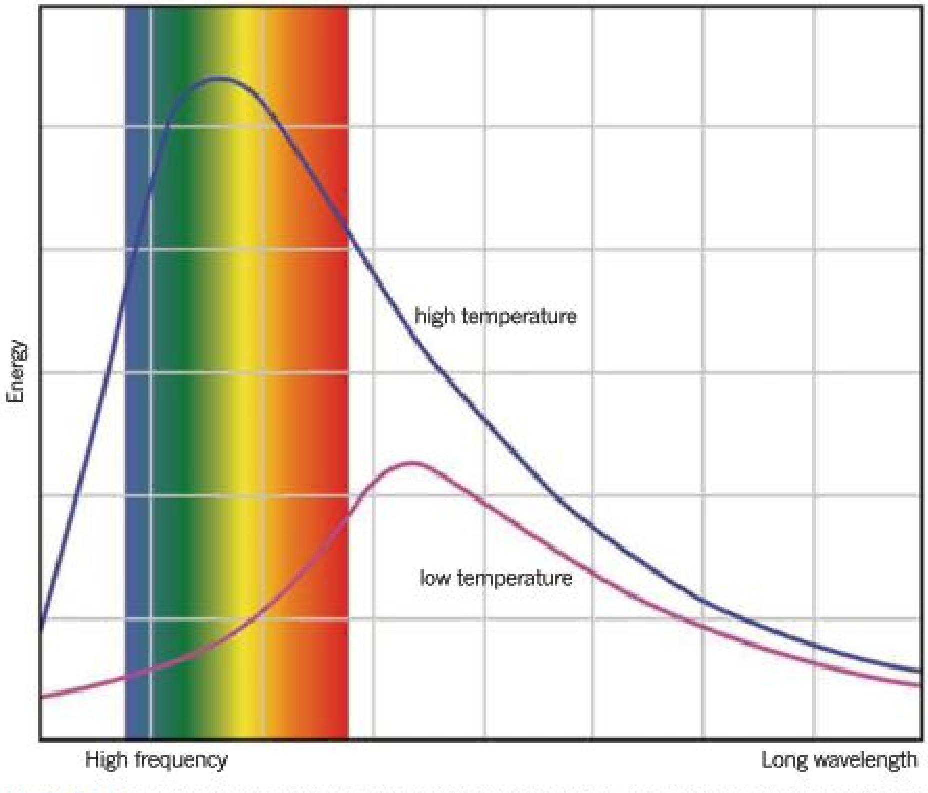


FIGURE 1.6.2 A system emits radiation over a range of frequencies. At a low temperature, it will emit small amounts of radiation of longer wavelengths. As the temperature of the system increases, more short-wavelength radiation is emitted and the total radiant energy emitted increases.

A human body emits radiation in the infrared range of wavelengths, while hotter objects emit radiation of a higher frequency and shorter wavelength. Hotter objects can emit radiation in the range of visible, ultraviolet and shorter wavelengths of the electromagnetic spectrum. For example, as a red-hot fire poker heats up further, it becomes yellow-hot.

EMISSION AND ABSORPTION OF RADIANT ENERGY

All objects both absorb and emit thermal energy by radiation. If an object absorbs more thermal energy than it emits, its temperature will increase. If an object emits more energy than it absorbs, its temperature will decrease. If no temperature change occurs, the object and its surroundings are in thermal equilibrium.

While all objects emit some radiation, they will not all emit or absorb at the same rate.

A number of factors affect both the rate of emission and the rate of absorption.

- Surface area—the larger the exposed surface area, the higher the rate of radiant transfer.
- Temperature—the greater the difference between the temperature of the absorbing or emitting surface and the temperature of the surrounding objects, the greater the rate of energy transfer by radiation.
- Wavelength of the **incident** radiation—matte black surfaces are almost perfect absorbers of radiant energy at all wavelengths. Highly reflective surfaces are good reflectors of all wavelengths. An example of how reflective surfaces can be exploited is shown in Figure 1.6.3. For all other surfaces, the absorption of particular wavelengths of radiant energy will be affected by the wavelength of that energy. For example, although white surfaces absorb visible wavelengths of radiant energy poorly, white surfaces will absorb infrared radiation just as well as black surfaces do.
- Surface colour and texture—the characteristics of the surface itself determine how readily that particular surface will emit or absorb radiant energy.

Matte black surfaces will absorb and emit radiant energy faster than shiny, white surfaces. This means that a roughened, dark surface will heat up faster than a shiny, light one. Matte black objects will also cool down faster since they will radiate energy just as efficiently as they absorb energy. Car radiators are painted black to increase the emission by radiation of the thermal energy that is collected from the car engine.

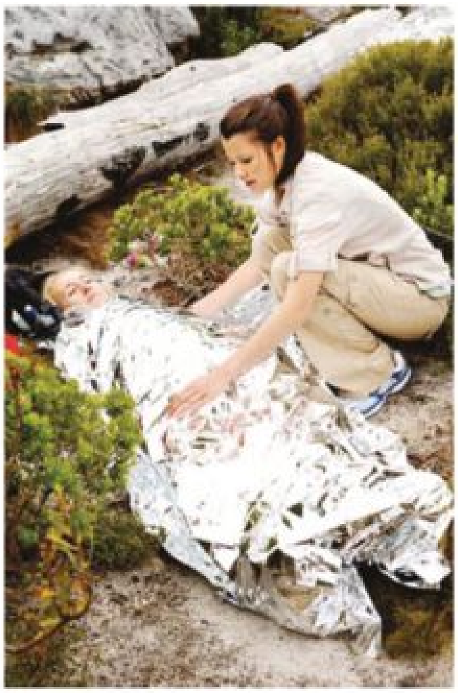


FIGURE 1.6.3 The silvered surface of an emergency blanket reflects thermal energy back to the body, and retains the radiant energy, which would normally be lost. This simple method works as excellent thermal insulation.

PHYSICSFILE

Heat loss in humans

It is estimated that, at normal room temperature, about 50 per cent of a person's heat loss is by radiation.

PHYSICSFILE

Thermal imaging

All objects emit radiant energy. Humans are warm-blooded and emit radiation in the infrared region of the electromagnetic spectrum. This radiation is not visible to us, but can be detected using thermal imaging devices or night-vision goggles. These devices are used by the military for surveillance, and by search and rescue personnel. Some animals, notably some varieties of bugs, are able to detect infrared radiation.

FIGURE 1.6.4 This person is difficult to see normally (a), but they are warmer than the trees surrounding them and produce a stronger infrared emission, which the thermal imaging device is able to detect (b).

1.6 Review

SUMMARY

- Any object whose temperature is greater than absolute zero emits thermal energy by radiation.
- Radiant transfer of thermal energy from one place to another occurs by means of electromagnetic waves.
- When electromagnetic radiation falls on an object, it will be partially reflected, partially transmitted and partially absorbed.
- The rate of emission or absorption of radiant heat will depend upon the:
  - temperature difference between the object and the surrounding environment
  - surface area and surface characteristics of the object
  - wavelength of the radiation.

KEY QUESTIONS

1 Light is shone on an object.

a List three interactions that can occur between the light and the object.

b Which of the interactions from part (a) are associated with a rise in temperature?

2 The wavelength and frequency of emitted radiation depends on the internal energy of an object. Complete the sentences below by choosing the correct option from those provided in the brackets.

The higher the temperature of the object, the [**higher**/**lower**] the frequency and the [**longer**/**shorter**] the wavelength of the radiation emitted. For example, if a particular object emits radiation in the visible range, a cooler object could emit light in the [**infrared**/**ultraviolet**] range of the electromagnetic spectrum.

3 Which of the following will affect the rate at which an object radiates thermal energy?

A its temperature
B its colour
C its surface nature (shiny or dull)
D none of these
E all of A–C

4 Why is it impossible for heat to travel from the Sun to the Earth by conduction or convection?

5 Thermal imaging technology can be used to locate people lost in the Australian bush. How can thermal imaging technology 'see' people when the naked eye cannot?

6 Three identical, sealed beakers are filled with near-boiling water. One beaker is painted matte black, one is dull white and the third is gloss white.

a Which beaker will cool fastest?

b Which beaker will cool slowest?

7 Computer chips generate a lot of thermal energy that must be dispersed for the computer to function efficiently. Devices called heat sinks are used to help this process. What would you predict the heat sinks to be made of?

Chapter review

01

KEY TERMS

absolute zero
absorption
conduction
conductor
convection
electromagnetic spectrum
emission
emit
evaporation
frequency
heat
incident
insulator
internal energy
kelvin
kinetic energy
kinetic particle model
latent heat
latent heat of fusion
latent heat of vaporisation
potential energy
radiation
specific heat capacity
temperature
thermal contact
thermal energy
thermal equilibrium
volatile
wavelength
work

1 According to the kinetic particle model, which of the following can be said of particles of matter?
- **A** They are in constant motion.
- **B** They have different sizes.
- **C** They have different shapes.
- **D** They are always floating.

2 What does the kelvin scale measure?

3 How does temperature differ from heat?

4 Convert:
- **a** 5°C to kelvin
- **b** 200 K to °C.

5 The specific heat capacity of iron is approximately half that of aluminium. A ball of iron and a ball of aluminium, both at 80°C, are dropped into a thermally insulated jar that contains a mass of water, equal to that of the balls, at 20°C. Thermal equilibrium is eventually reached. Describe the final temperatures of each of the metal balls.

6 Two cubes, one silver and one iron, have the same mass and temperature. A quantity Q of heat is removed from each cube. Which one of the following properties causes the final temperatures of the cubes to be different?
- **A** density
- **B** specific heat capacity
- **C** latent heat of vaporisation
- **D** volume

7 A solid substance is heated but its temperature does not change. Explain what is occurring.

8 Which possesses the greater internal energy—1 kg of water boiling at 100°C or 1 kg of steam at 100°C? Explain why.

9 A liquid is evaporating, causing the liquid to cool. Explain why the temperature of the liquid decreases.

10 A 2.00 kg metal object requires 5.02×10^3 J of heat to raise its temperature from 20.0°C to 40.0°C. What is the specific heat capacity of the metal in J kg^{-1} K^{-1}? Give your answer to the nearest whole number.

11 How many joules of energy are required to melt exactly 80 g of silver? ($L_{\text{fusion}} = 0.88 \times 10^5$ J kg^{-1}) Give your answer to two significant figures.

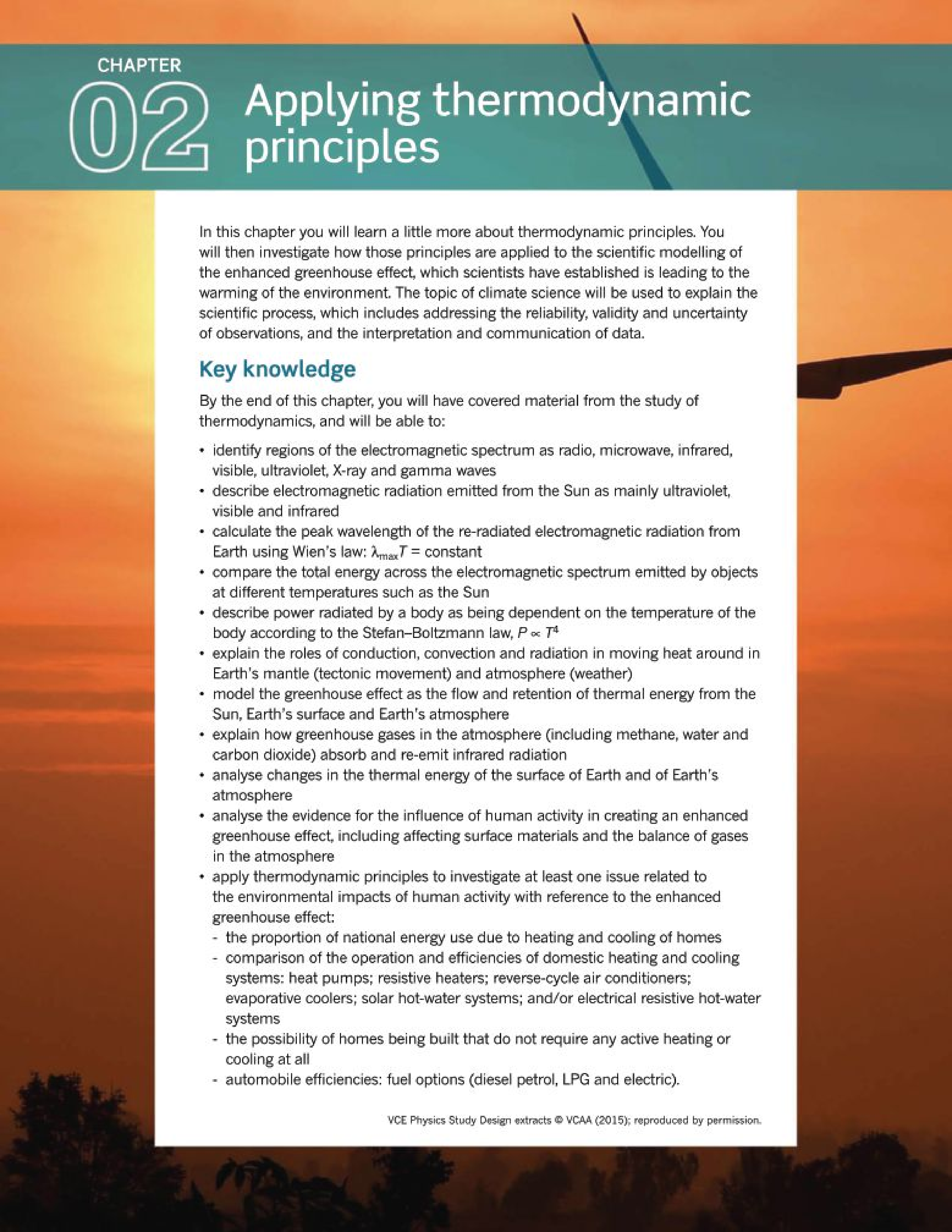
12 An insulated container holding 4.55 kg of ice at 0.00°C has 2.65 MJ of work done on it, while a heater provides 14600 J of heat to the ice. If the latent heat of fusion of ice is 3.34×10^5 J kg^{-1}, calculate the final temperature of the water. Assume that the increase in internal energy is first due to an increase in the potential energy and then an increase in the kinetic energy.

13 An ice-cube that is sitting on a block of polystyrene at room temperature (20°C) melts very slowly. An ice-cube sitting on a metal pan also at room temperature melts very rapidly. Explain why this is so and why the metal feels colder to touch, even though it is the same temperature as the polystyrene.

14 On a hot day when you sweat, your body feels cooler when a breeze is blowing. Explain why this happens.

15 A vacuum flask has a tight-fitting stopper at the top. Its walls are made up of an inner and outer layer, which are shiny and are separated by a layer of air. Describe how this design makes a vacuum flask good at keeping liquids hot inside.

16 Hypothermia is the cooling of the body to levels considerably lower than normal. The body's functions slow and death can result. A person may survive 12 hours in air at 0°C before suffering hypothermia, but may survive only a few minutes in water at 0°C. Explain why this is so and how wetsuits can help you survive in cold water.



CHAPTER

02 Applying thermodynamic principles

In this chapter you will learn a little more about thermodynamic principles. You will then investigate how those principles are applied to the scientific modelling of the enhanced greenhouse effect, which scientists have established is leading to the warming of the environment. The topic of climate science will be used to explain the scientific process, which includes addressing the reliability, validity and uncertainty of observations, and the interpretation and communication of data.

Key knowledge

By the end of this chapter, you will have covered material from the study of thermodynamics, and will be able to:

- identify regions of the electromagnetic spectrum as radio, microwave, infrared, visible, ultraviolet, X-ray and gamma waves
- describe electromagnetic radiation emitted from the Sun as mainly ultraviolet, visible and infrared
- calculate the peak wavelength of the re-radiated electromagnetic radiation from Earth using Wien's law: $\lambda_{max}T$ = constant
- compare the total energy across the electromagnetic spectrum emitted by objects at different temperatures such as the Sun
- describe power radiated by a body as being dependent on the temperature of the body according to the Stefan–Boltzmann law, $P \propto T^4$
- explain the roles of conduction, convection and radiation in moving heat around in Earth's mantle (tectonic movement) and atmosphere (weather)
- model the greenhouse effect as the flow and retention of thermal energy from the Sun, Earth's surface and Earth's atmosphere
- explain how greenhouse gases in the atmosphere (including methane, water and carbon dioxide) absorb and re-emit infrared radiation
- analyse changes in the thermal energy of the surface of Earth and of Earth's atmosphere
- analyse the evidence for the influence of human activity in creating an enhanced greenhouse effect, including affecting surface materials and the balance of gases in the atmosphere
- apply thermodynamic principles to investigate at least one issue related to the environmental impacts of human activity with reference to the enhanced greenhouse effect:
  - the proportion of national energy use due to heating and cooling of homes
  - comparison of the operation and efficiencies of domestic heating and cooling systems: heat pumps; resistive heaters; reverse-cycle air conditioners; evaporative coolers; solar hot-water systems; and/or electrical resistive hot-water systems
  - the possibility of homes being built that do not require any active heating or cooling at all
  - automobile efficiencies: fuel options (diesel petrol, LPG and electric).

2.1 Heating by radiation

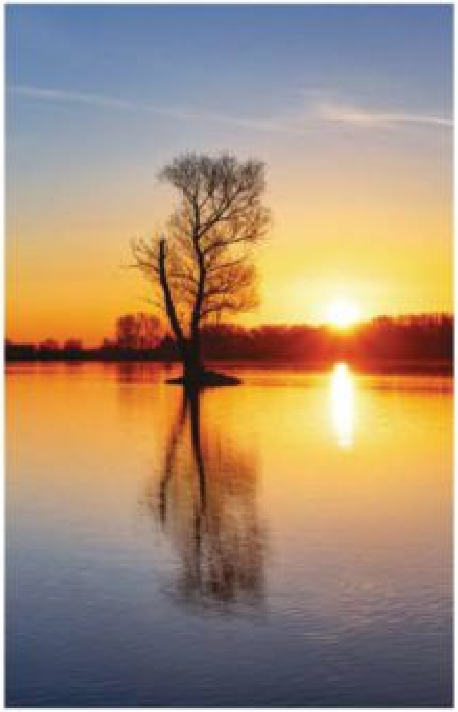


FIGURE 2.1.1 The energy from the Sun makes our world habitable and drives much of the Earth's climate.

Without the energy of the Sun, ecosystems like that shown in Figure 2.1.1 could not exist. All life on Earth depends upon the transfer of radiant energy from the Sun through space. The rate at which energy is lost from the Earth's atmosphere then determines how much the climate warms or cools. That rate will depend upon the rate at which thermal energy is reflected, retained or transmitted from the Earth. And, in turn, it is the composition of the atmosphere that determines the balances of those processes.

To understand the potential effects of human activity on the Earth's climate, it is important to understand and quantify how the Earth gains, circulates and loses thermal energy through these processes. This section looks at ideas behind quantifying (putting a numerical value on) the transfer of thermal energy by radiation.

THE ELECTROMAGNETIC SPECTRUM

Recall from Chapter 1 that in the context of thermal energy, the term **radiation** is a shortened form of **electromagnetic radiation** and includes such things as:

- visible light
- ultraviolet light
- infrared radiation.

The full range of electromagnetic radiation is called the electromagnetic spectrum.

The electromagnetic spectrum is divided into different categories according to the wavelength of the radiation and how the radiation is used. The different categories are illustrated in Figure 2.1.2. You will notice from the figure that infrared radiation overlaps with the lower end of the visible spectrum.

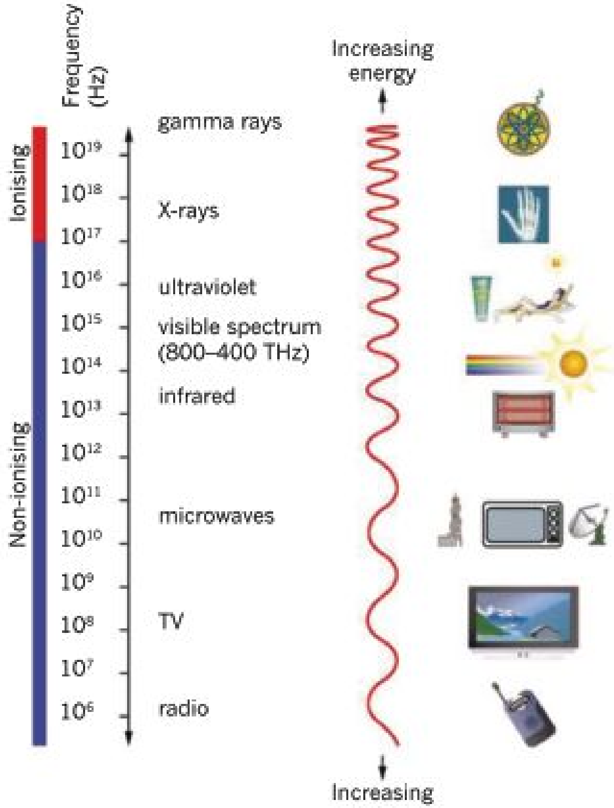


FIGURE 2.1.2 The electromagnetic spectrum showing the wavelength of the electromagnetic radiation and the corresponding frequencies (Hz). The spectrum is divided into categories based on wavelength and radiation production. Visible light is a mid-frequency range of wavelengths that can be seen by the human eye.

All forms of electromagnetic radiation are essentially the same, differing only in their frequency and, therefore, their wavelength.

The energy carried by electromagnetic radiation is proportional to the frequency. High-frequency gamma rays are at the high-energy end of the spectrum. Low-frequency radio waves carry the least energy and are at the low-energy end of the spectrum.

The seven categories of electromagnetic radiation are grouped according to similar characteristics. Table 2.1.1 shows these groups from the highest energy to the lowest. There is some overlap between these groups, largely based on application rather than distinct boundaries or values.

Gamma rays (γ-rays)	The highest energy, shortest wavelength energy is produced within the nucleus of an atom. Gamma rays are one of the three types of emissions that come from radioactive (unstable) atoms.
X-rays	When fast-moving electrons are fired into an atom, X-rays are produced. X-rays got their name as a result of scientists at first not knowing what they were, hence the letter 'X'.
Ultraviolet (UV)	UV light has a shorter wavelength than visible violet light and is less energetic than gamma rays or X-rays but is known to cause skin cancer, particularly with frequent exposure. Wavelengths are less than 10 nm (1 nm = 10^{-9} m).
Visible light	This is the small portion of wavelengths around the middle of the electromagnetic spectrum that can be detected by human eyes. Many other life forms, for example insects and birds, can perceive wavelengths well into the ultraviolet range.
Infrared (IR)	Infrared or heat radiation is emitted by all objects that are not at a temperature of absolute zero. The hotter the object, the more radiation emitted and the shorter the wavelength within the IR band.
Microwaves	The microwaves that cook your dinner and allow remote communications and radar to work are produced by the spin of electrons or nuclei. Wavelengths range from about 1 nm to 10 cm.
Radio and television waves	Electrons oscillating in a conducting wire, such as an antenna, produce the radio and television waves that bring music and pictures to your home and carry voice and data to your phone. Long wavelength, low-energy electromagnetic radio and television waves can be transmitted across very long distances.

TABLE 2.1.1 Types of electromagnetic radiation.

PHYSICSFILE

Scientific notation

Very small and very large numbers, such as the wavelengths and frequencies of electromagnetic radiation, can be tedious to write out in full. Scientific notation, which uses powers of 10, is a useful way of showing numbers in a more concise way. Table 2.1.2 lists some commonly used powers of 10 factors, prefixes and symbols.

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{12}	tera	T	10^{-3}	milli	m
10^{9}	giga	G	10^{-6}	micro	μ
10^{6}	mega	M	10^{-9}	nano	n
10^{3}	kilo	k	10^{-12}	pico	p

TABLE 2.1.2 Common prefixes, symbols and factors for large and small numbers.

WIEN'S LAW

Electromagnetic radiation is emitted by all objects and systems whose temperature is above absolute zero (0 K or −273°C). The actual wavelength or frequency of the emitted radiation depends almost entirely on the internal energy of the object and not on the characteristics of the material itself.

The higher the temperature of an object or system, the higher the frequency and the shorter the wavelength of the emitted radiation. As temperature increases, electromagnetic radiation is emitted at increasingly higher frequencies. Consider the following examples.

- Cool objects, such as the human body, emit radiation at long wavelengths with lower energy, such as infrared radiation. Infrared radiation is not visible under normal circumstances.
- At higher temperatures, objects emit radiation with a higher frequency and you can see it glow red. An example is a bar heater that glows red hot.
- At even higher temperatures, say 2000 K, objects such as the filament of an incandescent light glow yellow or white.
- Very hot objects, at temperatures of 10^6 K or more, emit the majority of their radiation within the gamma and X-ray regions of the electromagnetic spectrum.

Wilhelm Wien, a German physicist, formulated laws that describe the properties of heat radiation. Wien discovered that the peak wavelength at which an object will emit the maximum intensity of radiation is dependent on its surface temperature. Wien's displacement law, more commonly known just as Wien's law, can be used to determine the peak wavelength for an object at a particular surface temperature.

Wien's law states that

$$\lambda_{max}T = 2.898 \times 10^{-3} \text{ m K}$$

where λ_{max} is the peak wavelength of the emitted radiation in metres (m)

T is the surface temperature of the object in kelvin (K).

That is, no matter what the surface temperature of an object, the product of the temperature and the wavelength at which the peak intensity of the emitted radiation occurs is a constant and is equal to about 2.898×10^{-3} m K.

The graph in Figure 2.1.3 shows the continuous spectrum emitted by any solid, liquid or even a dense gas at particular temperatures. An object at a temperature of around 12 000 K will emit its peak wavelength in the ultraviolet range. That is, the wavelength corresponding to the highest intensity for the 12 000 K curve will occur in the ultraviolet range.

The 6000 K curve in the graph in Figure 2.1.3 corresponds to the surface temperature of our Sun. The intensity maximum corresponds to a peak wavelength within the visible band of the electromagnetic spectrum, at about 500 nm.

The surface temperature of the Sun means that the electromagnetic radiation emitted by the Sun peaks in the visible range and is emitted mainly within the range between ultraviolet and infrared, including visible light.

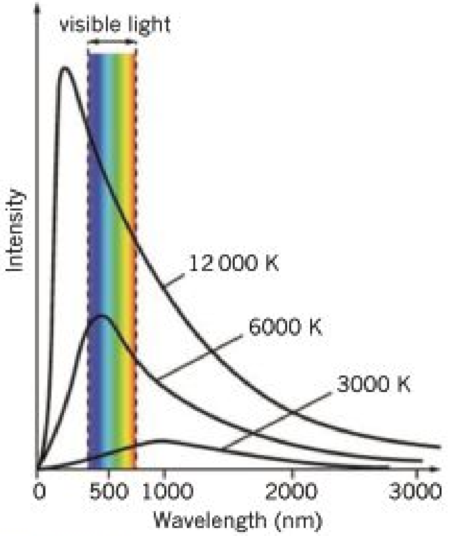


FIGURE 2.1.3 The spectrum of wavelengths emitted for an idealised black body at different temperatures. The radiation approximates the surface temperature of many real objects.

PHYSICSFILE

Our eyes and the Sun

It is no coincidence that the human eye is very good at detecting visible light. Human eyes have evolved to be most receptive to wavelengths of light within what is known as the visible range and which correspond to the highest intensity of light produced by our Sun. If the Sun had a lower surface temperature, of say around 3000 K, it's highly probable that human eyes would be adapted to the infrared range.

EXTENSION

Black-body radiation

Wien's work on the wavelength of the radiation emitted by a hot, dense object was based initially on a theoretical object called a **black body**. A black body does not necessarily have to be black. An incandescent lamp or a light bulb may be, to a certain extent, regarded as a black body.

This theoretical object completely absorbs all the rays of electromagnetic radiation incident on it regardless of the wavelength of the radiation. In other words, a black body does not reflect any radiation. The radiation emitted by many objects, such as the Sun, can be approximated as the radiation emitted by a black body at the same temperature. The spectrum emitted by a hot solid, liquid or dense gas is continuous but has a peak intensity at a wavelength inversely proportional to the surface temperature. This relationship is more simply stated by rearranging Wien's law:

$\lambda_{\text{max}} \propto \frac{1}{T}$ and also $T \propto \frac{1}{\lambda_{\text{max}}}$

Wien's law makes it possible to determine the approximate temperature of stars, assuming that they emit radiation similar to that emitted by a black body. During astronomic observations, it was discovered that stars at different temperatures have peaks in the graph of emissive power at different wavelengths. When the wavelength that corresponds to the peak of the power emitted by a star is known, the temperature of the star can be found by applying Wien's law.

Wien's law can be used to find the temperature of an object, including our Sun.

Worked example 2.1.1

THE TEMPERATURE AT A STAR'S SURFACE

The Sun emits a continuous electromagnetic spectrum with a peak wavelength of approximately 500 nm. Based on this wavelength, estimate the surface temperature of the Sun.

Thinking	Working
Express the peak wavelength in metres.	$\lambda_{\text{max}} = 500\text{ nm} = 500 \times 10^{-9}\text{ m}$
Rearrange Wien's law to solve for T.	$\lambda_{\text{max}}T = 2.898 \times 10^{-3}\text{ m K}$ $T = \frac{2.898 \times 10^{-3}}{\lambda_{\text{max}}}$
Substitute the value for λ_{max} and solve for T.	$T = \frac{2.898 \times 10^{-3}}{500 \times 10^{-9}}$ $= 6000\text{ K}$

Worked example: Try yourself 2.1.1

THE TEMPERATURE AT A STAR'S SURFACE

A newly discovered star is observed to have a peak emitted radiation wavelength of approximately 90 nm. Based on this wavelength, estimate the surface temperature of this star.

PHYSICSFILE

Wilhelm Wien

Wilhelm Wien (1864–1928) was a German physicist. Wien was awarded the 1911 Nobel Prize in Physics for making a significant contribution to the thermodynamics of radiation. In 1893 he had discovered Wien's displacement law, which paved the way for Planck's quantum theory of radiation since Wien's law was only valid at high frequencies. Planck underestimated the radiance at low frequencies, but later corrected the theory and proposed what is now called Planck's law. This led to the development of quantum theory, the theoretical basis of modern physics, which explains the nature and behaviour of matter and energy. So Wien's displacement law was a very significant discovery indeed!

FIGURE 2.1.4 German physicist Wilhelm Wien (1864–1928).

PHYSICS IN ACTION

Warm white or cool daylight

The simple act of replacing a light globe requires many considerations. The globe needs to be the right size and have the appropriate power rating. Further consideration needs to be given to the technology—a standard incandescent, a hot halogen, low-energy fluorescent or LED. Finally, there's the choice of 'warm white' or 'cool daylight'. Some of the different types of globe available are shown in Figure 2.1.5.

Warm white and cool daylight are descriptions used by manufacturers to describe the colour of the light from what is basically a white light globe. A warm white globe provides a slightly yellow, 'warm' glow, while 'cool daylight' is a more harsh white light. This difference relates to the peak wavelength emitted by the globe. Manufacturers design globes to mimic surface temperatures corresponding to particular peak wavelengths. Warm white is usually labelled as corresponding to a surface temperature of 3000 K; cool daylight corresponds to 6500 K.

Computer manufacturers use similar colour profiles to allow computer users to correct the whiteness of a computer screen.

FIGURE 2.1.5 Standard, energy saving and LED light bulbs come in various shades of white that relate to the surface temperature they mimic.

Re-radiated electromagnetic radiation

Radiant energy interacts with matter in three ways. It can be:

- reflected
- transmitted
- absorbed.

More often than not it will be a combination of two or more of these modes. For example, absorption of radiant energy by the surface of the Earth leads to the heating of the surface and re-radiation of radiant energy.

The Earth's atmospheric gases absorb very different narrow wavelengths of the incoming solar radiation depending upon the nature of the gas. The smaller molecules of oxygen and nitrogen absorb very short wavelengths of solar radiation.

The larger molecules of water vapour and carbon dioxide absorb primarily longer infrared radiant energy. About 17% of the radiant energy from the Sun is absorbed by the atmosphere, leading to the heating of the upper layers of the atmosphere.

The lower layers of the atmosphere are not actually predominately heated by the Sun's radiant energy but, rather, by radiation from the Earth. The Earth is much cooler than the Sun, and so emits much longer wavelength radiation. With the temperature of the Sun's surface being approximately 6000 K, the peak wavelength of solar radiation is around 500 nm. This corresponds to the visible part of the electromagnetic spectrum. The Earth has an average temperature of around 16°C or 289 K. At that significantly lower temperature, the Earth emits most of its energy in the infrared range of the electromagnetic spectrum. This can be seen in the infrared image shown in Figure 2.1.6.

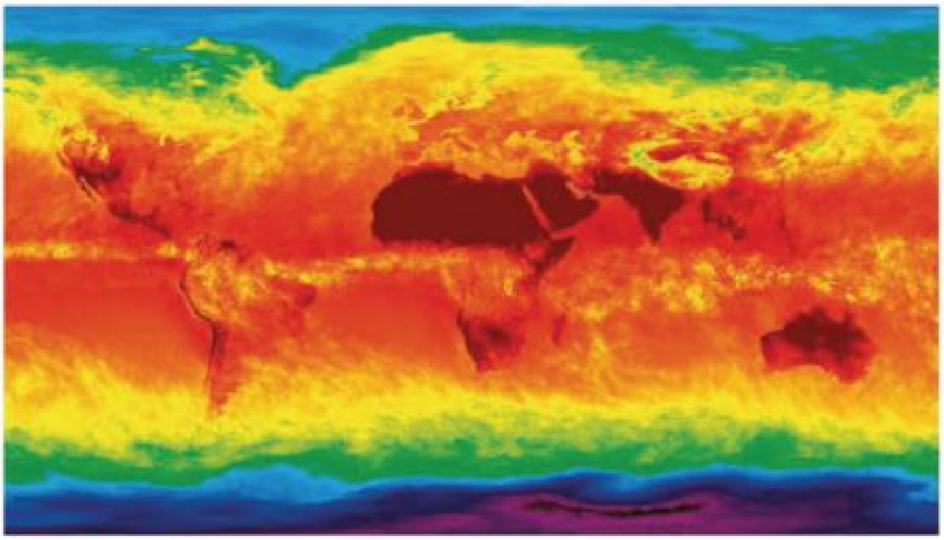


FIGURE 2.1.6 The Atmospheric Infrared Sounder (AIRS) instrument aboard NASA's Aqua satellite senses temperature using infrared wavelengths. This image shows the temperature of the Earth's surface or clouds covering it for the month of April 2003.

Using Wien's law, the peak wavelength of re-radiated energy from the Earth can be calculated.

Worked example 2.1.2

RE-RADIATED ENERGY FROM THE EARTH

The Earth's average surface temperature is 289 K. What is the peak wavelength of the re-radiated electromagnetic radiation?

Thinking	Working
State Wien's law.	$\lambda_{max}T = 2.898 \times 10^{-3}$ m K
Rearrange Wien's law to express it in terms of λ_{max}.	$\lambda_{max} = \frac{2.898 \times 10^{-3}}{T}$
Substitute the value for T and solve for λ_{max}.	$\lambda_{max} = \frac{2.898 \times 10^{-3}}{289}$ $= 1.00 \times 10^{-5}$ m = 10 μm

Worked example: Try yourself 2.1.2

RE-RADIATED ENERGY FROM THE EARTH

The Earth's average surface temperature at the equator is 300 K. What is the peak wavelength of the re-radiated electromagnetic radiation from this portion of the Earth?

EMITTED AND ABSORBED ENERGY

While all objects above absolute zero emit some radiation, they will not all emit energy at the same rate. There are a number of factors that will affect the total rate of energy emission by radiation across the electromagnetic spectrum—surface area, temperature, and surface colour and texture. An object's effectiveness at emitting energy is called its **emissivity** and is given the symbol *e*. Emissivity is defined as the fraction of energy that is emitted relative to that emitted by a thermally black surface (a black body). A black body is a perfect emitter of heat energy and has an emissivity value of 1. A material with an emissivity value of 0 would be considered a perfect reflector of thermal energy. Good emitters are also good absorbers.

As an example, matte black surfaces will emit radiant energy at a greater rate than shiny, white surfaces. Matte black surfaces have a value of *e* close to 1, whereas shiny surfaces have an *e* value close to 0. This means that a roughened, dark surface will heat up faster than a shiny, light one. It will also cool down faster since it will radiate energy just as efficiently. Car radiators are painted black for this reason—to increase the emission of thermal energy collected from the car engine.

Table 2.1.3 shows the emissivity of some common materials.

Material	Emissivity
anodised aluminium	0.77
brick	0.81–0.86
concrete	0.92
powdered charcoal	0.96
glass	0.92
ice	0.97
black paper	0.90
black rubber stopper	0.97
human skin	0.98
snow	0.80
water	0.95

TABLE 2.1.3 The emissivity of some common materials.

STEFAN–BOLTZMANN EQUATION

Ludwig Boltzmann and Joseph Stefan, two Austrian physicists, discovered the basic laws that govern heat radiation in 1879 and put these factors together. They found that the rate at which an object radiates energy (i.e. its power) is related to the temperature in kelvin of the object.

$$P \propto T^4$$

where P is the power in watts (W)

T is the temperature of the object in kelvin (K).

That is, an object at 2000 K will emit 2^4 or 16 times as much energy via radiation compared with one at 1000 K.

Worked example 2.1.3

RATE OF ENERGY RADIATION BY PROPORTION

The surface temperature of a particular object increases from 500 K to 2000 K. What effect does this have on the rate of transmission of thermal energy from the object?

Thinking	Working
Find the factor by which the temperature has increased or decreased.	$T_{\text{multiple}} = \frac{T_{\text{final}}}{T_{\text{initial}}}$ $= \frac{2000}{500} = 4$
Use the relationship to find the effect on thermal energy transmission.	$P \propto T^4$ $P \propto 4^4$ $P \propto 256$ The rate of transmission is increased by two hundred and fifty-six times.

Worked example: Try yourself 2.1.3

RATE OF ENERGY RADIATION BY PROPORTION

The surface temperature of a particular object increases from 900 K to 1800 K. What effect does this have on the rate of transmission of thermal energy from the object?

The graph in Figure 2.1.7 shows the rate at which radiant energy per unit area is emitted (the radiated emittance) for objects at different temperatures. The graph allows a direct comparison of the energy being emitted across the electromagnetic spectrum. Note that the peak wavelength shown by this measure is the same as that for the earlier graph of intensity versus wavelength shown in Figure 2.1.3 on page 32.

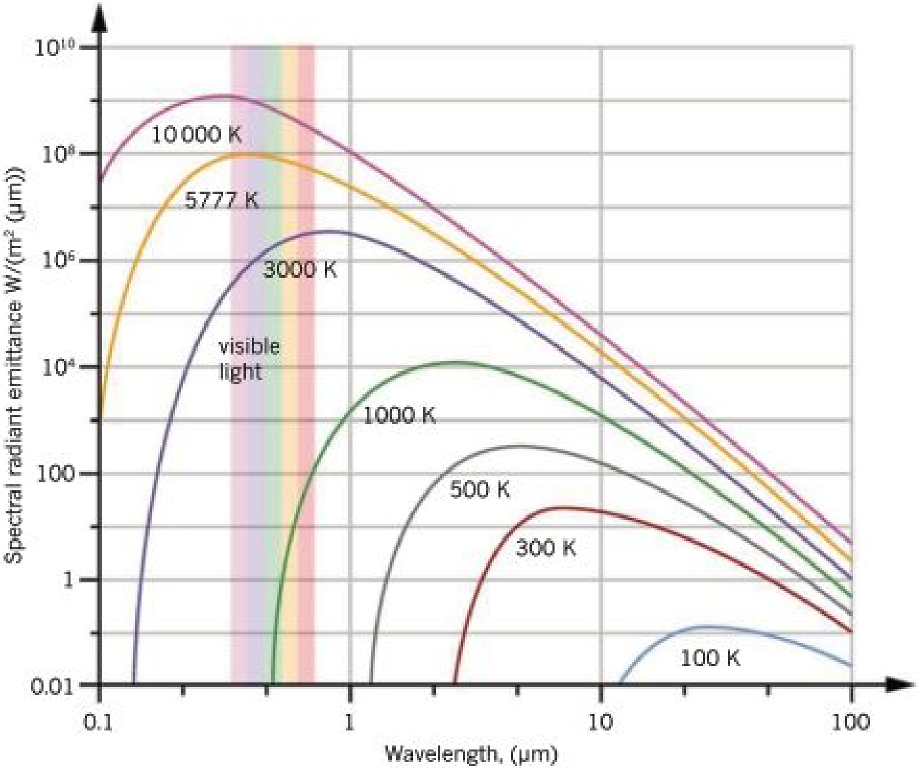


FIGURE 2.1.7 A graph of radiated emittance for objects at a range of temperatures. The approximate surface temperatures of the Sun (5777 K) and the Earth (300 K) are shown for comparison.

The Stefan–Boltzmann relationship can be expressed more precisely by considering the effect of other variables on the rate of transfer of radiant energy:

ℹ $P = e\sigma AT^4$

where P = power in watts (W)

T is the temperature of the object in kelvin (K)

A is the surface area in metres squared (m^2)

σ is the a universal constant called the Stefan–Boltzmann constant

$= 5.67 \times 10^{-8}$ W m^{-2} K^{-4}

e is the the emissivity of the surface.

This equation is called the Stefan–Boltzmann equation after its developers.

If the object is at a temperature T and the temperature of the surroundings is T_s, then:

ℹ $P = e\sigma A(T^4 - T_s^{\,4})$

Note that in the equation above, if P is negative, this means that the surroundings are warmer than the object and heat is transferred to the object rather than from the object.

PHYSICSFILE

Solar radiation

The Earth receives 1.8×10^{17} J of radiant energy from the Sun every second. That radiant energy is distributed across the electromagnetic spectrum with most of it lying within or close to the visible spectrum. When the Sun is directly overhead, each square metre of the Earth's surface receives an average of 1365 J per second. At the spring and autumn equinox when the Sun is directly above the equator, on a relatively clear day, cities around the latitudes of Australia's southern state capitals will still receive about 1000 J per second per square metre.

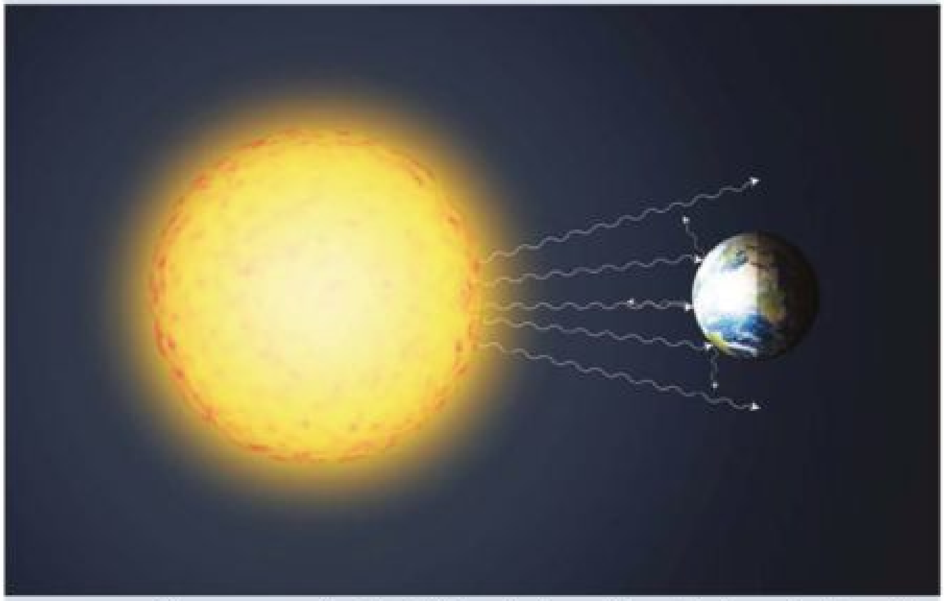


FIGURE 2.1.8 More energy reaches the Earth from the Sun each hour than is used by the world in one year.

Worked example 2.1.4

RATE OF ENERGY RADIATION BY CALCULATION

A black plastic water tank ($e = 0.95$) holds 1000 L of hot water at 90°C. The tank is a cube of 1 m length on each side. Estimate the rate of heat loss from the tank, assuming the surrounding environment is at 20°C.

Thinking	Working
Convert temperature from degrees Celsius to kelvin.	$T = 90°\text{C} = 363\text{ K}$ $= 20°\text{C} = 293\text{ K}$
Work out the surface area of the cube.	Surface area of a cube = 6 × area of one side $A = 6 \times 1 \times 1 = 6\text{ m}^2$
Substitute values into the Stefan–Boltzmann equation to find the rate of energy transmission, P.	$P = e\sigma A(T^4 - T_s^4)$ $= 0.95 \times 5.67 \times 10^{-8} \times 6 \times (363^4 - 293^4)$ $= 3230\text{ W}$

Worked example: Try yourself 2.1.4

RATE OF ENERGY RADIATION BY CALCULATION

A galvanised water tank ($e = 0.28$) holds 1000 L of hot water at 90°C. The tank is a cube of 1 m length on each side. Estimate the rate of heat loss from the tank, assuming the surrounding environment is at 20°C.

2.1 Review

SUMMARY

- In the context of thermal energy, the term radiation is a shortened form of electromagnetic radiation. The full range of electromagnetic radiation is called the electromagnetic spectrum.
- The seven categories of the electromagnetic spectrum, from highest energy, shortest wavelength to lowest energy, longest wavelength, are: gamma rays (γ-rays), X-rays, ultraviolet (UV), visible light, infrared (IR), microwaves and radio waves.
- The peak wavelength, at which an object will emit the maximum intensity of radiation, is dependent on the object's surface temperature and is given by Wien's law. Wien's law states that $\lambda_{max}T = 2.898 \times 10^{-3}$ m K.
- The surface temperature of the Sun, approximately 6000 K, means that the electromagnetic radiation emitted by the Sun peaks in the visible range and is emitted mainly within the range between ultraviolet and infrared light.
- The average surface temperature of the Earth is approximately 289 K. Re-radiated energy from the Earth will lie in the infrared section of the electromagnetic spectrum. The peak wavelength can be calculated using Wien's law.
- A graph of radiated emittance, in W m^{-2}, for objects at different temperatures can be used to compare the relative emittance of different objects at different temperatures across the electromagnetic spectrum.
- The rate at which an object radiates energy (its power) is proportional to the fourth power of the temperature (in kelvin) of the object: $P \propto T^4$
- The rate of transfer of radiant energy can be calculated using the Stefan–Boltzmann equation: $P = e\sigma AT^4$
- If the object is at a temperature T and the temperature of the surroundings is T_s, then: $P = e\sigma A(T^4 - T_s^4)$

KEY QUESTIONS

1 If the wavelength of radiation from ice is approximately 10^{-5} m, in what region of the electromagnetic spectrum does it lie?

2 If a star emits a continuous electromagnetic spectrum with a peak wavelength of approximately 800 nm, what is the surface temperature of the star?

3 The element of an electric heater is just seen to glow a dull red. This colour corresponds to the lower end of the visible spectrum at approximately 700 nm. What temperature, in kelvin, is the element of the heater?

4 If a star has a surface temperature of 9000 K, what is the peak wavelength of the energy being emitted by the star?

5 The surface temperature of an electric stove top doubles. What will happen to the rate of transmission of thermal energy?

A It will stay the same.
B It will double.
C It will increase 8 times.
D It will increase 16 times.

6 The emissivity of a person's head when covered with hair is approximately $e = 0.75$. A bald head has an emissivity of approximately $e = 0.25$ at a similar temperature. What will the rate of absorption of radiant thermal energy from the Sun by a bald head be, compared to a head that's full of hair?

7 A black rubber ball of radius 12 cm has an emissivity of 0.98. How much power is radiated by the ball when it is at a temperature of 25°C? Give your answer to two significant figures.

8 A person is sitting, unclothed, in a room which is at a temperature of 15°C. Calculate the rate at which thermal energy is lost by radiation if the person's skin temperature is 34°C and $e = 0.70$. Assume that the exposed area of the person's skin is 1.5 m^2.

9 The Earth receives on average about 390 W m^{-2} of radiant thermal energy from the Sun, averaged over the whole of the Earth. It radiates an equal amount back into space, maintaining a thermal equilibrium that keeps the average temperature on Earth the same. Assuming the Earth is a perfect emitter of radiant energy ($e = 1$), estimate the average surface temperature of the Earth in °C.

2.2 The enhanced greenhouse effect

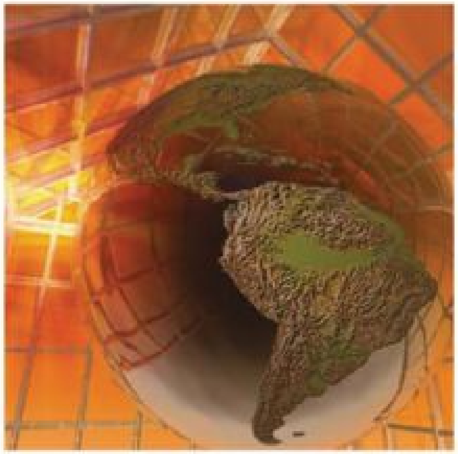


FIGURE 2.2.1 A conceptual illustration of the greenhouse effect, with the Earth warming up in a giant greenhouse.

Earth's atmosphere has long acted as a greenhouse, trapping some of the Sun's energy to keep the Earth at a constant temperature. You can imagine the Earth as being inside a giant greenhouse, as is shown in Figure 2.2.1. Without this greenhouse effect, the Earth would be a very cold place.

In recent years, evidence has shown increasing levels of specific gases in the atmosphere, in particular carbon dioxide.

These gases are responsible for most of the absorption of longer wavelength radiant energy. As the levels of these gases rise, more thermal energy is being retained than previously. This has led to the idea of an 'enhanced' greenhouse effect.

It's the recent enhancement of this effect that is the cause for concern among climate scientists. This section looks at the concepts behind the theory of the greenhouse effect and how thermal energy is moved around the Earth.

HEATING THE EARTH

The overall temperature of the Earth is determined by the total thermal energy received and the amount that is lost back to space. Any change in that balance will lead to a warming or cooling of the Earth as a whole.

Energy gain—radiant energy from the Sun

The vast majority of the thermal energy received by the Earth is short-wave radiant energy from the Sun. Most of this is within or close to the visible spectrum.

Of the incoming radiant energy from the Sun, about 47% is eventually absorbed by the Earth's surface. As shown in Figure 2.2.2, of the remaining energy:

- about 23% of the energy reaching the Earth is absorbed by the atmosphere, predominantly by the ozone layer, but also by water vapour and greenhouse gases such as carbon monoxide and carbon dioxide
- about 26% is reflected back towards outer space by clouds
- about 4% is reflected by the Earth's surface.

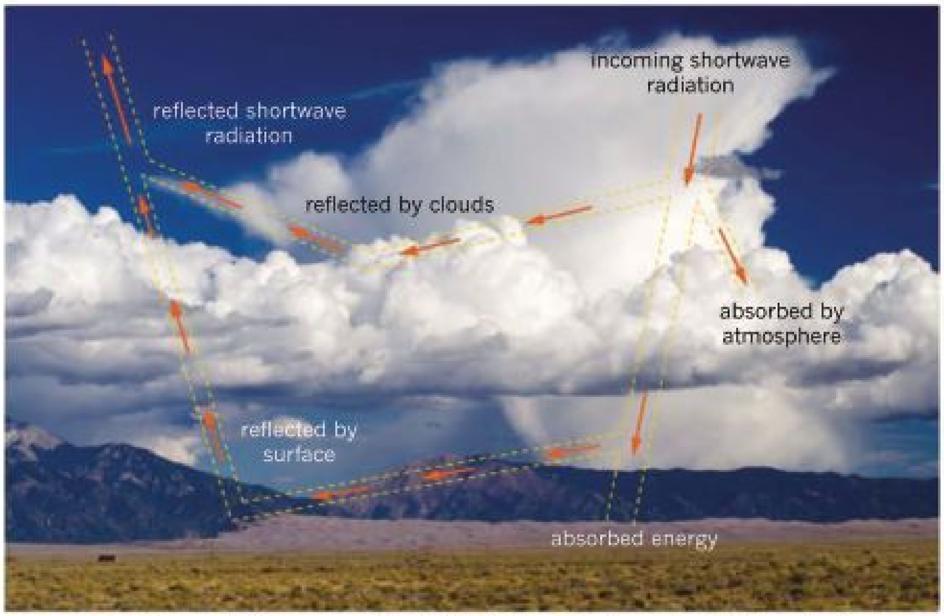


FIGURE 2.2.2 Of the radiant energy reaching the Earth from the Sun, only 47% is eventually absorbed by the Earth's surface. The remainder is reflected back into space or absorbed by the atmosphere.

The radiant energy which is absorbed by the Earth's surface transforms to thermal energy, increasing the temperature of the absorbing surfaces and of the air in contact with those surfaces. Over the longer term very little of this energy is retained. Almost all is re-radiated by the surface as long-wavelength radiant energy.

Energy retention—greenhouse gases

With the Earth's surface re-radiating heat from the Sun back out towards space, you may wonder how the Earth remains warm. The Earth stays warm because the **greenhouse gases** in the atmosphere absorb some of this energy and re-radiate it down towards Earth's surface again. Only about 5% is lost directly out to space. This energy retention has led to a long-term energy balance, allowing life to evolve.

The long wavelength infrared radiation emitted by Earth is readily absorbed by these gases in the atmosphere. This creates a desirable sort of **greenhouse effect** and the relatively consistent day and night temperatures required by the life forms that have evolved on Earth.

This cycle is shown in Figure 2.2.3.

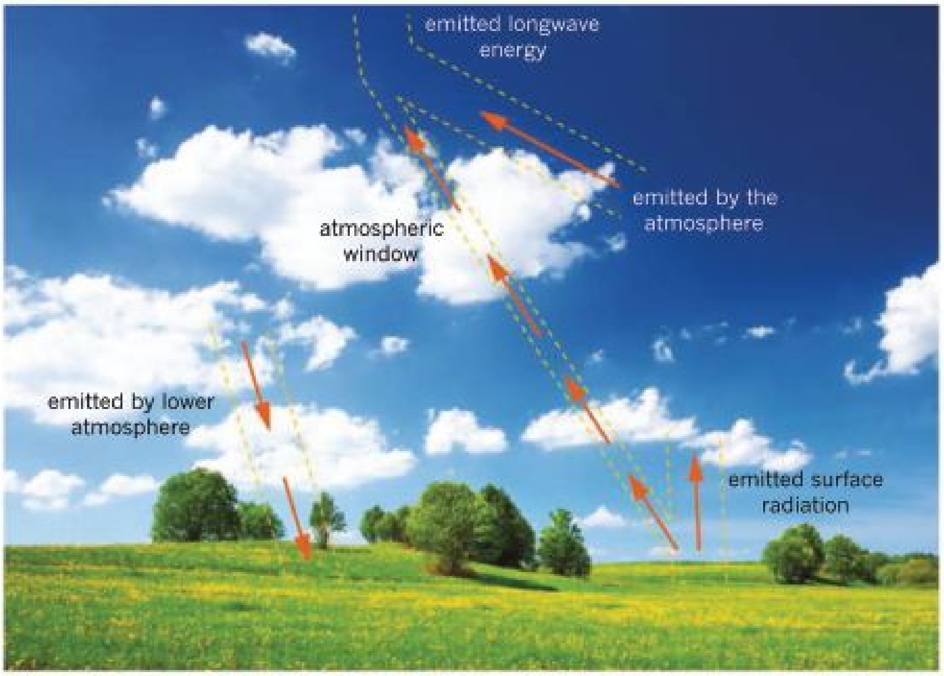


FIGURE 2.2.3 The Sun's energy that has been absorbed by the Earth is re-radiated away from the Earth as infrared radiation. Significant amounts of this are absorbed by the greenhouse gases in the atmosphere and are re-radiated back to the Earth's surface.

ENHANCED GREENHOUSE EFFECT

In recent years, evidence has suggested that the atmosphere is absorbing and retaining more of the long-wavelength infrared radiation emitted from the Earth's surface. This is called the **'enhanced' greenhouse effect**.

Specific gases present in the lower atmosphere in very small proportions are responsible for most of the absorption of these longer wavelengths. The most significant of these have increased in concentration over the last 100 years, coinciding with the industrialisation of modern society. These greenhouse gases are gases that can absorb and emit long-wavelength infrared radiation.

The most abundant gases in the Earth's atmosphere are, in order of abundance:

- water vapour (H_2O)
- carbon dioxide (CO_2)
- methane (CH_4)
- nitrous oxide (N_2O)
- ozone (O_3)
- chlorofluorocarbons (CFCs).

The major atmospheric gases (nitrogen, oxygen and argon) are almost totally unaffected by infrared radiation and so have no effect on the absorption or re-emission of infrared wavelengths.

PHYSICSFILE

Venus greenhouse effect

Many scientists believe that Venus used to have an environment similar to that on Earth, with lower temperatures and even liquid water on the surface. Now, the carbon dioxide atmosphere on Venus is 92 times denser than Earth's atmosphere at the surface. The average surface temperature on Venus is 462°C. It is thought that Venus experienced a 'runaway' greenhouse effect.

On Venus, approximately 10% of the Sun's energy reaches the surface and heats it up. The surface radiates the infrared energy back towards space. As you can see in the image below, this radiant energy used to continue out to space, but it is now retained due to the dense atmosphere, which has caused the planet to heat up.

FIGURE 2.2.4 The greenhouse effect on Venus. The changed atmosphere, which now consists mainly of carbon dioxide, blocks most of the radiated heat energy from leaving the atmosphere.

The enhanced greenhouse effect is caused by the combined effects of each greenhouse gas. The extent to which each gas contributes depends on the chemical characteristics of the gas, on its percentage abundance and on some indirect effects, such as water vapour turning to ice. For example, while methane absorbs 72 times as much thermal energy as carbon dioxide, it is present in much smaller concentrations so overall it doesn't contribute as much to the enhanced greenhouse effect. Table 2.2.1 ranks greenhouse gases based on their overall contribution to the enhanced greenhouse effect.

Compound	Formula	Contribution (%)
water vapour/clouds	H_2O	36–72%
carbon dioxide	CO_2	9–26%
methane	CH_4	4–9%
ozone	O_3	3–7%
nitrous oxide	N_2O	1.5%
chlorofluorocarbons	CFCs	0.1%

TABLE 2.2.1 Greenhouse gas contribution to the enhanced greenhouse effect.

Aside from CFCs, which are entirely human produced, most greenhouse gases have natural sources as well as those from human activity. The sources of some greenhouse gases are summarised in Table 2.2.2.

Gas	Natural sources	Human sources
carbon dioxide, CO_2	respiration volcanic eruptions	burning fossil fuels deforestation land-use changes
methane, CH_4	digestion in animals	decomposition of wastes in landfills agriculture and especially rice cultivation energy use domestication of livestock
nitrous oxide, N_2O	soils under natural vegetation and the oceans	fertiliser use burning fossil fuels nitric acid production biomass burning
chlorofluorocarbons (CFCs)	none	industrial processes refrigerants (such as those used in air-conditioning) a variety of consumer products

TABLE 2.2.2 Sources of greenhouse gases.

Greenhouse gas levels

By comparing current levels of greenhouse gases with those found trapped in air in Antarctic ice cores, scientists are able to develop a good idea of the changes in greenhouse gas levels over long periods of time. Prior to the Industrial Revolution (which started in the middle of the 1700s), concentrations of these gases were relatively constant. In the modern industrial era, human activities have increased the proportions of greenhouse gases in the atmosphere. This has mainly occurred through the burning of **fossil fuels** and large-scale land clearing. The graph in Figure 2.2.5 clearly shows how significant the rise in greenhouse gases has been in modern times.

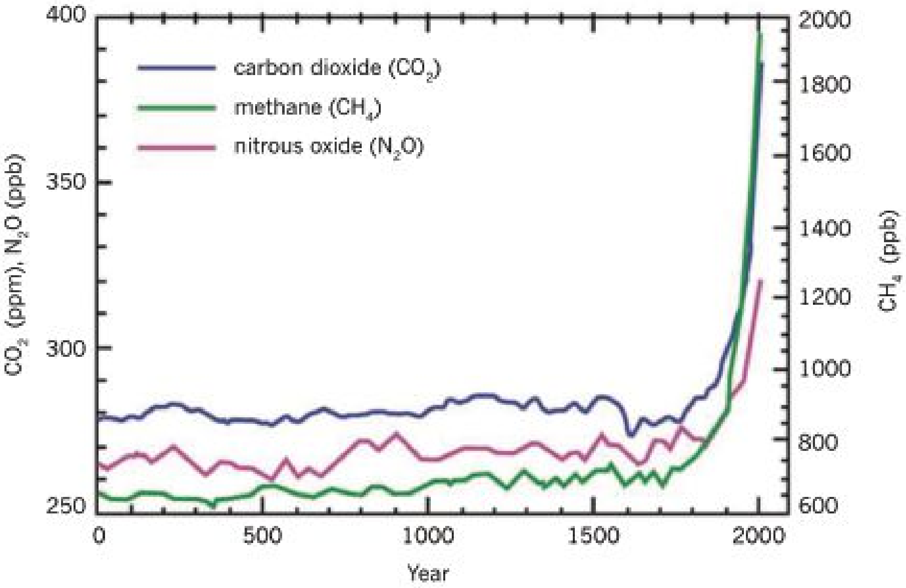


FIGURE 2.2.5 This graph shows that atmospheric carbon dioxide, methane and nitrous oxide have increased since the Industrial Revolution.

For example, based on data from ice-core samples, carbon dioxide was found to have been relatively constant at between 180 and 300 parts per million (ppm) for the last 400 000 years, up until the Industrial Revolution. Figure 2.2.6 shows the levels of carbon dioxide (CO_2), specifically in the atmosphere. The graph shows that in 2014, carbon dioxide levels in the northern hemisphere passed 400 ppm. With increased levels of carbon dioxide comes an increase in the absorption of infrared radiation and the re-radiation back to the Earth's surface rather than reflection into space.

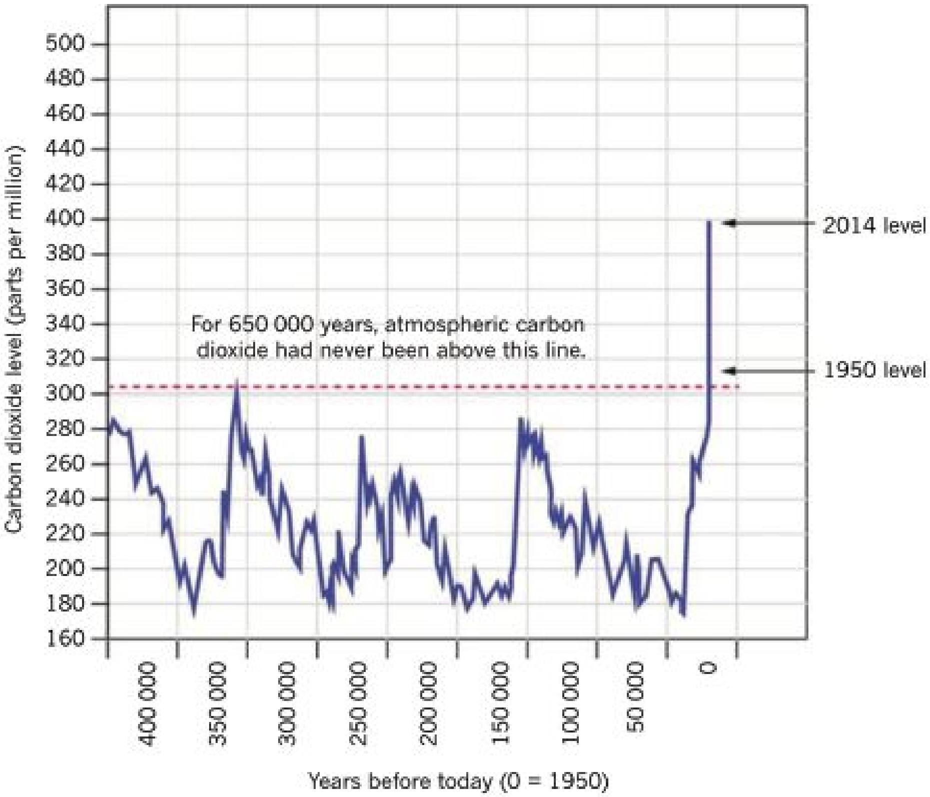


FIGURE 2.2.6 Carbon dioxide levels over the last 400 000 years, based on ice-core samples. The recent sudden rise coincides with the beginning of the industrial age.

Carbon dioxide levels are not the only gas levels that have increased, as you can see in Table 2.2.3.

Greenhouse gas	Pre-industrial-age concentration	Current concentration
carbon dioxide, CO_2	280 ppm (parts per million)	400 ppm
methane, CH_4	722 ppb (parts per billion)	1800 ppb
nitrous oxide, N_2O	270 ppb	325 ppb
ozone, O_3	237 ppb	337 ppb

TABLE 2.2.3 Changes in greenhouse gas concentrations over the industrial era.

Historical data based on ice-core samples lets us be quite confident of these changing greenhouse gas levels.

PHYSICS IN ACTION

Rising carbon dioxide levels

As levels of carbon dioxide in the environment rise, the enhanced greenhouse effect increases and the Earth gets warmer and warmer. The consequences for the environment can be devastating. Currently, the Earth's average temperature (land and ocean) is about 15°C, which is an increase of about 0.8°C since 1880. If the Earth keeps getting warmer, it will impact the entire planet: the oceans, weather patterns and all living things. If the Earth becomes too hot, it is likely to cause significant changes which will prevent the Earth from sustaining our current way of life. Such changes include reducing the Earth's capacity to grow enough food. There is evidence that these changes are occurring already. The diagrams in Figure 2.2.7 show what could happen if carbon dioxide levels continue to change and therefore cause the average global temperature to change.

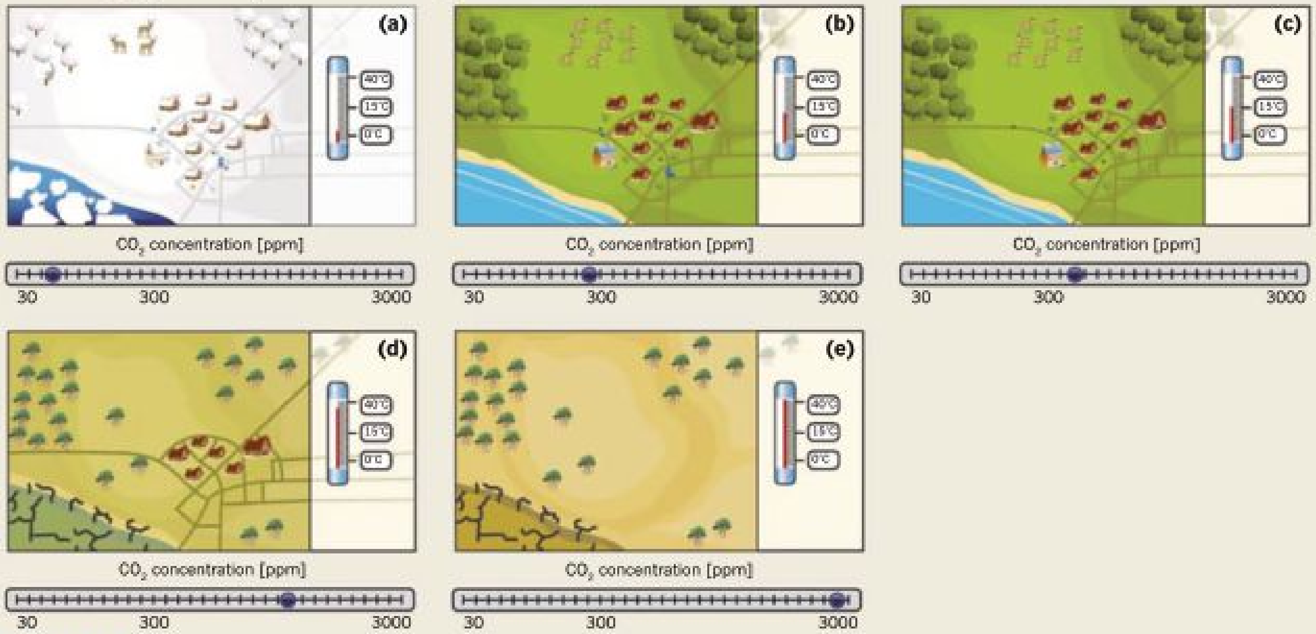


FIGURE 2.2.7
(a) With CO_2 levels very low, the Earth would be far colder than it is now. Average global temperatures would be around 0°C. Some greenhouse gases are needed to maintain a balance and keep the Earth warm enough to sustain life.
(b) At 300 parts per million (ppm) of CO_2, the Earth's average temperature is below 15°C.
(c) Earth's CO_2 levels have recently reached 400 ppm.
(d) The Earth will become drier if the CO_2 levels and therefore the average global temperature increases.
(e) If it gets too hot, the Earth will not be able to sustain our current way of life.

HUMAN ACTIVITIES AND ENERGY RE-RADIATED BY THE EARTH

As you know, very little of the radiant energy that reaches the Earth's surface from the Sun is retained. However, the rate at which energy is re-radiated by the Earth's surface depends on the surface material. Recall from Section 2.1 that materials with a high emissivity will not only readily absorb thermal energy, but will also re-radiate the energy very well.

Energy retention: Building

The building of cities has affected the greenhouse effect in a number of ways.

- The materials used to build cities have a high emissivity, meaning they increase the rate at which thermal energy is re-radiated into the atmosphere.
- Increased concentrations of greenhouse gases around cities can also act as urban heat traps. This leads to increased localised temperatures and more thermal energy in cities than rural areas.

Studies by the bureaus of meteorology around the world have found that the centre of a modern city may be several degrees warmer than the surrounding suburbs and countryside. The Melbourne Central Business District (CBD) may be as much as 7°C warmer than outlying suburbs. This can be seen in the thermal image in Figure 2.2.8. While that may sound great in the depths of winter, in mid-summer it becomes a cause of greater heat-related mortality (deaths), damage to infrastructure and a factor in increased energy use as people rely on air-conditioning to stay cool.

FIGURE 2.2.8 In this thermal image of Melbourne, you can see the bluer park areas (such as Flemington racecourse, top left, or the MCG, bottom right) are cooler than roads, which crisscross the image in red.

Much of this additional energy comes from the dark surfaces, like bitumen roads, thermal energy waste by vehicles and poorly-insulated buildings. The effect is compounded by more carbon dioxide in the air from car exhausts, building heating and so on. Dark surfaces release energy overnight, keeping nighttime temperatures in cities 2°C or so above those of surrounding suburbs.

Energy retention: Land clearing

It has been estimated that about 70% of Australia's original native vegetation has been cleared in the 200 years since European settlement. The majority of that clearing has happened in the last 50 years. The map in Figure 2.2.9 shows the areas most affected by clearing.

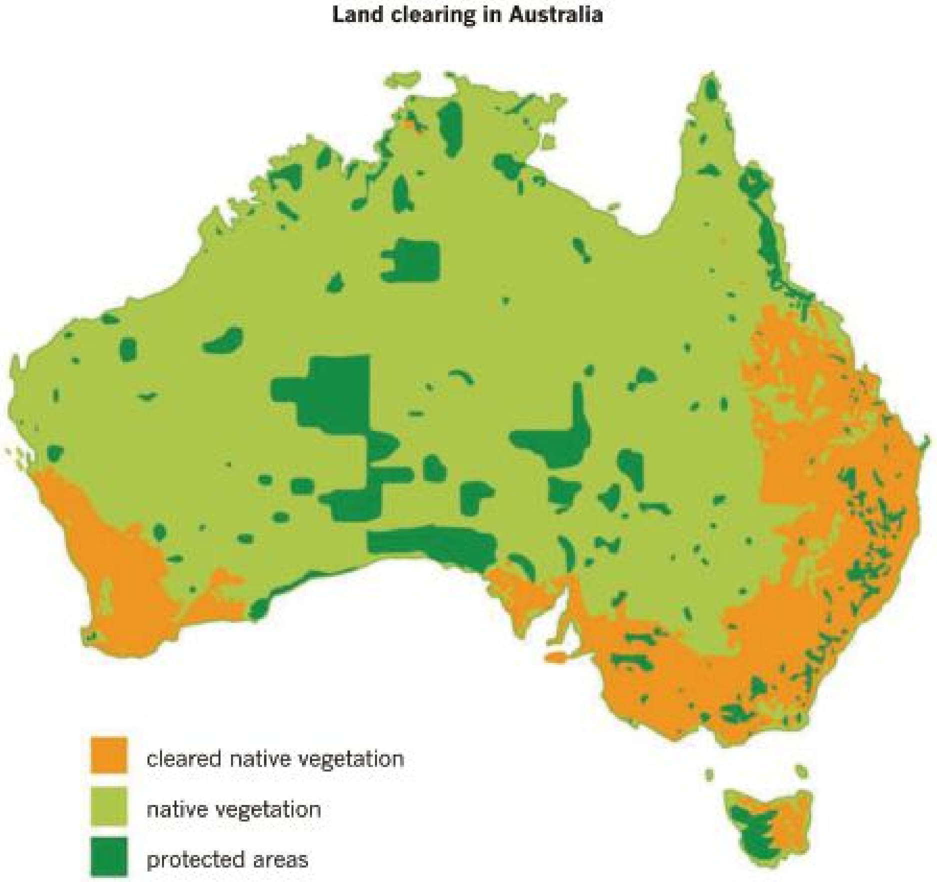


FIGURE 2.2.9 Land cleared in Australia since European settlement.

Land clearing on a large scale changes the amount of thermal energy retained or reflected by changing the surface characteristics. In Australia, land clearing contributes approximately 12% of Australia's total emissions. Clearing vegetation causes:

- an increase in the number of dry days
- an increase in days over 35°C
- a decrease in daily rainfall intensity and cumulative rainfall on rainy days
- an increase in the duration of droughts.

Clearing native vegetation also contributes to higher temperatures and decreased rainfall by reducing both shade and humidity.

MOVING HEAT AROUND THE EARTH

Scientists build climate models that can predict the effects the enhanced greenhouse effect is having, or will have, on the climate. This requires an understanding of the processes by which thermal energy is moved around the Earth.

The main mechanisms for moving heat around the Earth are conduction, convection and radiation. Evaporation also moves heat energy around the Earth's atmosphere and is sometimes considered a fourth process. More correctly though, evaporation is a product of heat transfer due to conduction, convection and radiation. (These processes are described in more detail in Chapter 1.)

Heat flow inside the Earth

While the vast majority of the thermal energy needed to support life on Earth comes from the Sun, there is a small proportion that comes from the Earth itself as **geothermal energy**.

At the Earth's core, it is estimated that temperatures are the same or higher than that on the surface of the Sun, at around 7000 K.

Heat flow inside the Earth is mainly through convection. Heat from the **mantle** in the centre of the Earth moves towards the surface, as shown in Figure 2.2.10. Heat flows constantly from within the Earth to the Earth's surface at a rate estimated to be around 47 terawatts (4.7×10^{13} W). That sounds like an enormous amount, but it is an average of 0.087 W m^{-2} or just 0.03% of the total radiant energy from the Sun that is absorbed by the Earth.

When the hot magma reaches the surface it transfers heat to the crust through conduction and then sinks back to the centre of the Earth. Where the crust is thinnest, conduction is higher, as shown in the thermal image in Figure 2.2.11.

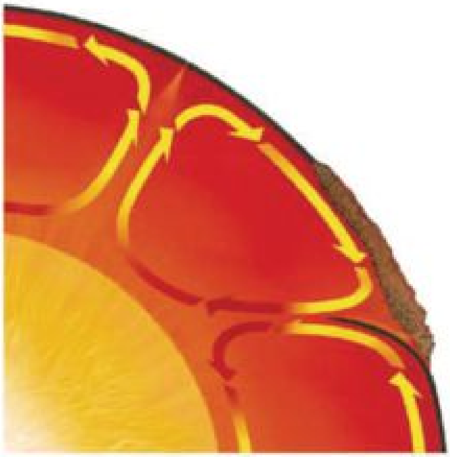


FIGURE 2.2.10 Mantle convection in the modern Earth. Hot plumes reach the outer layer of the upper mantle and fan out before finally sinking as cooler magma, having transferred thermal energy through the Earth's crust by conduction.

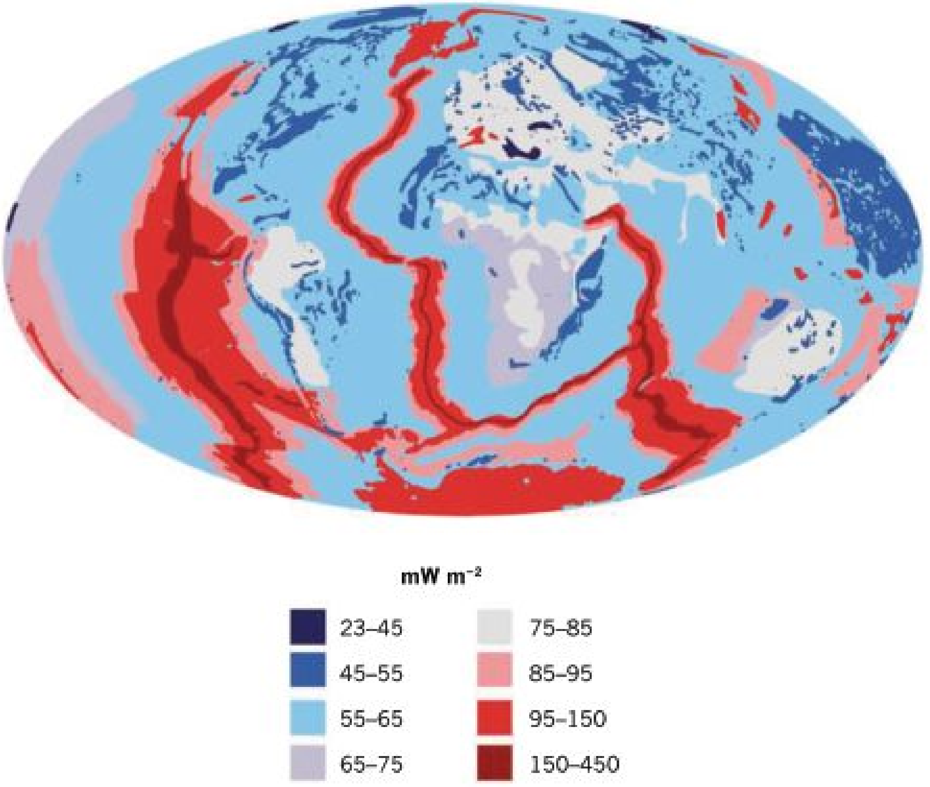


FIGURE 2.2.11 The flow of heat by conduction from Earth's interior to the surface. Higher heat exchanges are observed at the locations of mid-ocean fault lines, where the crust is thinnest.

Heat flow in the oceans

Ocean currents are mass movers of thermal energy over very large distances through the process of convection. The large-scale circulation of water and thermal energy via ocean currents is called the great ocean conveyor. The conveyor occurs because of both variations in water temperature and salinity (salt concentration).

While the direction of ocean currents is not due to convection, warm water in the ocean will rise and cold water will sink through normal convection. Tropical regions of the Earth's oceans receive a large amount of radiant energy from the Sun. This warmed water travels via currents towards the polar regions. The water cools at the poles, sinks and begins its journey back to the equator. Currents carrying warm, less-dense water move on the surface in one direction while cold, salty water moves in the opposite direction.

The major surface currents of the Earth's oceans are caused by the wind and are influenced by land masses. The prevailing winds of the Earth's atmosphere push the surface water along until it reaches land, at which point the currents will divert along the coasts of the land masses. In the major ocean basins, surface currents form circular patterns which are influenced by the Earth's rotation. Currents in the northern hemisphere flow clockwise, and in the southern hemisphere flow anticlockwise.

The Gulf Stream is part of the great ocean conveyor and is shown in the centre of Figure 2.2.12. The Gulf Stream carries the warm, salty water up along the east coast of the Americas, then towards Europe. It makes the climate of Western Europe much warmer than other regions at the same latitudes. At colder northern latitudes, the water becomes so dense that it sinks to the sea floor and travels south (shown in blue).

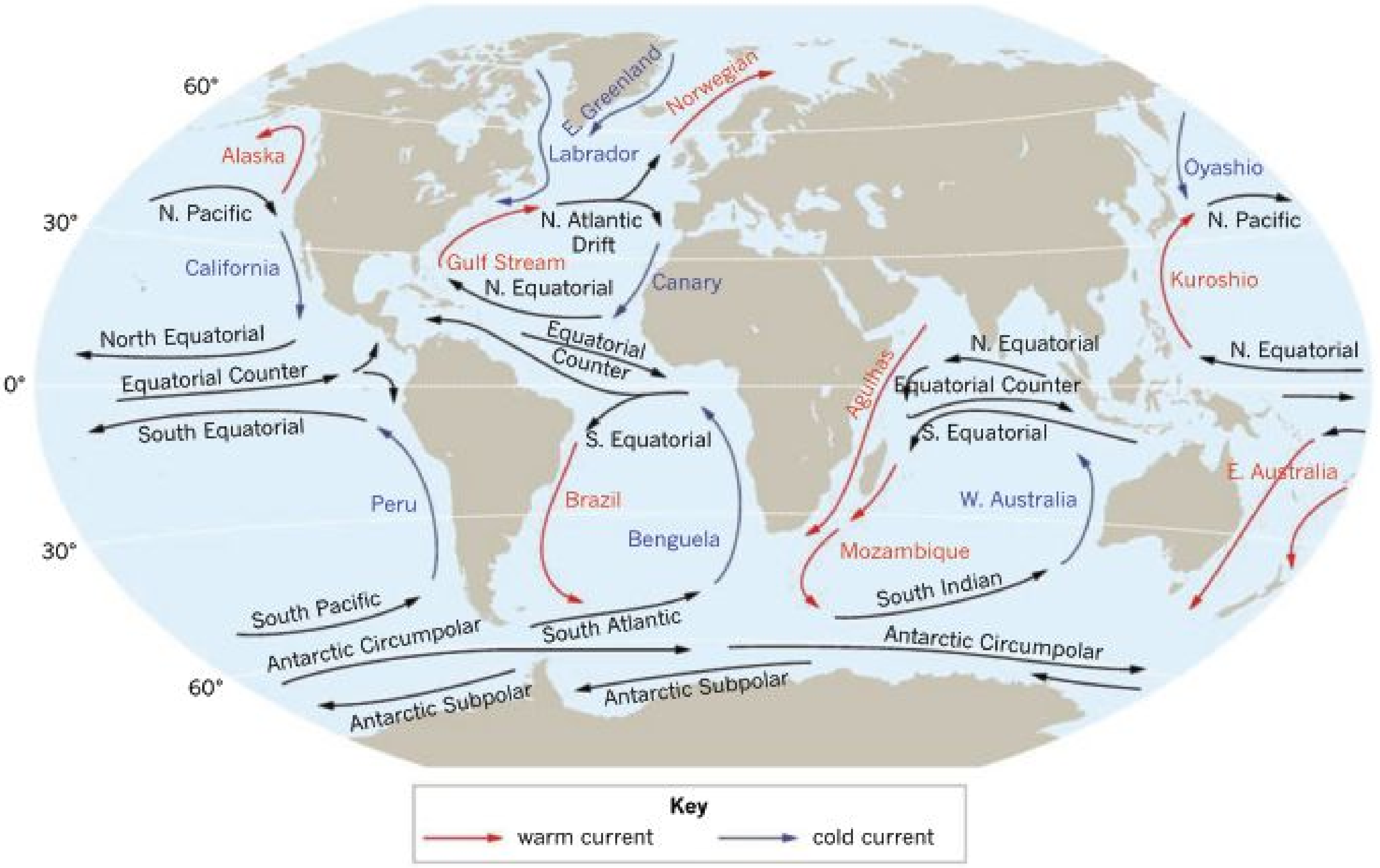


FIGURE 2.2.12 The major surface currents of the Earth's oceans. Warm currents are shown in red, cold currents in blue.

The Earth's present system of ocean currents is one of delicate balance. Minor changes can upset the balance of temperature differentials and prevailing winds. Scientists modelling the Earth's climate have become concerned at the significant melting of Greenland's ice caps. There is also concern about large-scale melting of ice in the Antarctic. Large-scale melting produces large volumes of low-density fresh water. This could prevent the surface currents in the north from sinking and returning as a southwards deep current. In turn, this would cause oceans and regions closer to the equator to warm while northern Europe would get colder.

Circulation via this process is extremely slow, taking around 1600 years to complete one cycle. While slow, this system of ocean currents plays a major role in determining the climate of many of the regions of the Earth.

PHYSICS IN ACTION

The Panamanic Seaway

The Panamanic Seaway, or Central American Seaway, was an ancient body of water separating North and South America. It formed around 200–150 million years ago during the separation of the **Pangaea** supercontinent. It allowed the flow of warm water from the Pacific Ocean through to the Atlantic Ocean. The open seaway is shown in Figure 2.2.13(b). Volcanic activity 2.5 million years ago caused the Panamanian **isthmus** to form, closing the seaway.

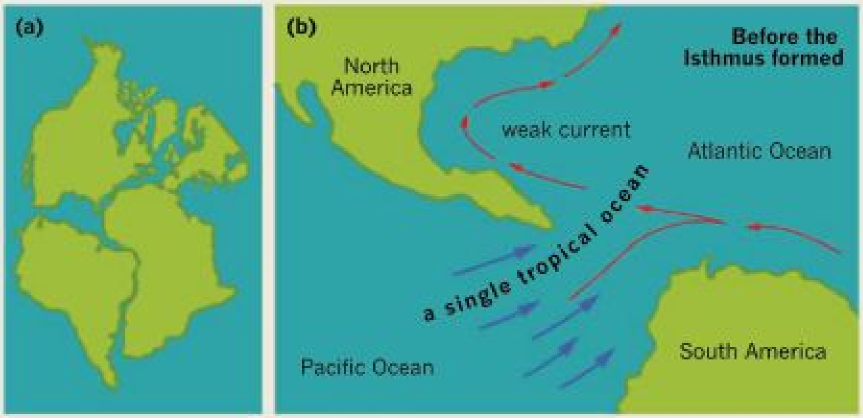


FIGURE 2.2.13 (a) A simplified sketch of the western part of the Pangea supercontinent. (b) The ancient seaway between North and South America was called the Panamanic Seaway.

The closure of the Panamanic Seaway had an enormous effect on ocean currents and global temperatures. The warm equatorial currents that had previously existed were closed off. Cold Arctic and Antarctic waters then lowered the temperature of the isolated Atlantic Ocean, bringing on an Ice Age. Climates became cooler and drier and ice sheets grew in Antarctica; glaciers formed in New Zealand and Tasmania.

Heat flow in the atmosphere

The Earth's atmosphere is an essential part of the Earth's climate system. It is able to absorb and store thermal energy, so it acts as a heat sink and has a major impact on our atmosphere. By transferring thermal energy quickly around the Earth, it is our atmosphere that regulates the temperature of the Earth and keeps it stable.

Energy is transferred within the atmosphere by:

- radiation—radiant energy largely comes from the Sun as short-wavelength radiation and, to a lesser degree, long-wavelength reflection and emission from the Earth's surface. Greenhouse gases retain radiant energy within the atmosphere.
- conduction—happens only in the very low levels of the Earth's atmosphere as air is a very poor conductor of thermal energy.
- convection—the major process by which thermal energy is moved around the Earth's atmosphere.

At the coast there is often a temperature difference between the land and the sea. The water in the sea hardly changes temperature between night and day due to its high specific heat capacity (covered in Chapter 1), but the land can become much hotter through the day. As the air is heated, it rises and is replaced by cooler, denser air from over the sea. Figure 2.2.14 shows how the cycle works.

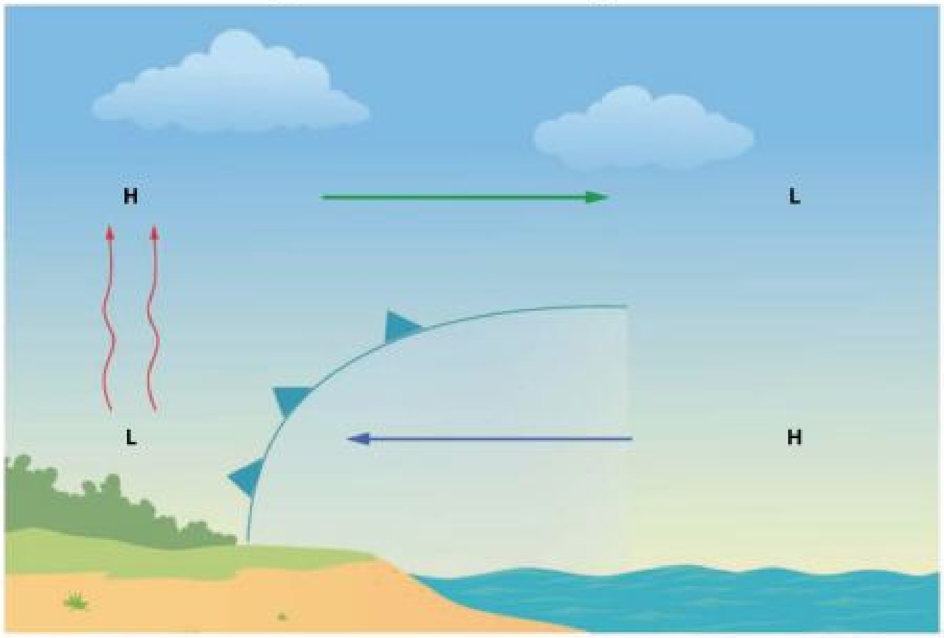


FIGURE 2.2.14 Sea breezes are created by convection currents caused by the temperature difference between air over the sea and air over the land. In this diagram, L represents regions of relatively low pressure, while H shows regions of higher pressure.

This moving air creates a sea breeze and is experienced in most coastal areas in Australia during summer. This makes the coastal climate more pleasant on a hot day than inland regions.

During the night the process is reversed, as shown in Figure 2.2.15. The land and the air above it cools more quickly. This cooler, denser air moves out over the ocean, displacing the now relatively lighter and warmer air, and a land breeze is created.

FIGURE 2.2.15 The convection currents created at night are the opposite of those during the day. At night, the land cools more quickly, denser air moves over the ocean and a land breeze occurs.

On a global scale, the radiant energy from the Sun at the equator heats the air and it becomes less dense. Cooler, denser air moves in and forces the warm air higher into the atmosphere. This creates an area of low pressure. Once the warm air is high in the atmosphere, it spreads out towards the poles, where it cools down. The cooler air sinks back to the Earth's surface creating areas of high pressure. These circular currents are created purely by convection and are one of the main transport mechanisms of thermal energy in the atmosphere.

PHYSICS IN ACTION

The Earth's oceans as a temperature buffer

Water has a significantly higher specific heat capacity than air. As such, the world's oceans provide a significant temperature buffer and climate stabilising effect, absorbing extra thermal energy when colder than the atmosphere and releasing it when warmer. This is evident in differences in weather and temperature between the northern and southern hemispheres.

The Earth's land masses and oceans are not distributed evenly. You can see this by looking at the world map in Figure 2.2.16. The northern hemisphere is approximately 61% ocean and 39% land. The southern hemisphere is approximately 81% ocean and just 19% land. The larger proportion of water in the southern hemisphere means that the average temperature variation between summer and winter in the southern hemisphere is 7.3°C. In the northern hemisphere the temperature variation is up to 14.3°C.

FIGURE 2.2.16 The difference in proportion of land and sea between the southern and northern hemispheres is immediately apparent, when you look at (composite) satellite images.

EXTENSION

Australia's climate

The modelling of the influence of the enhanced greenhouse effect on Australia's climate is affected by two other major climate phenomena:

- the Southern Oscillation
- the Indian Ocean Dipole.

Southern Oscillation

The Southern Oscillation is a sequence of changes to the way the atmosphere and water circulate across the Pacific Ocean. Changes to the Southern Oscillation have significant effects on the climate of the countries across the tropical regions of the Pacific Ocean, including Australia.

At one extreme of the Southern Oscillation is an El Niño event. This event causes drier conditions in eastern Australia, often leading to droughts. On the other side of the Pacific, South America experiences warmer, wetter conditions. The El Niño event is illustrated in Figure 2.2.17.

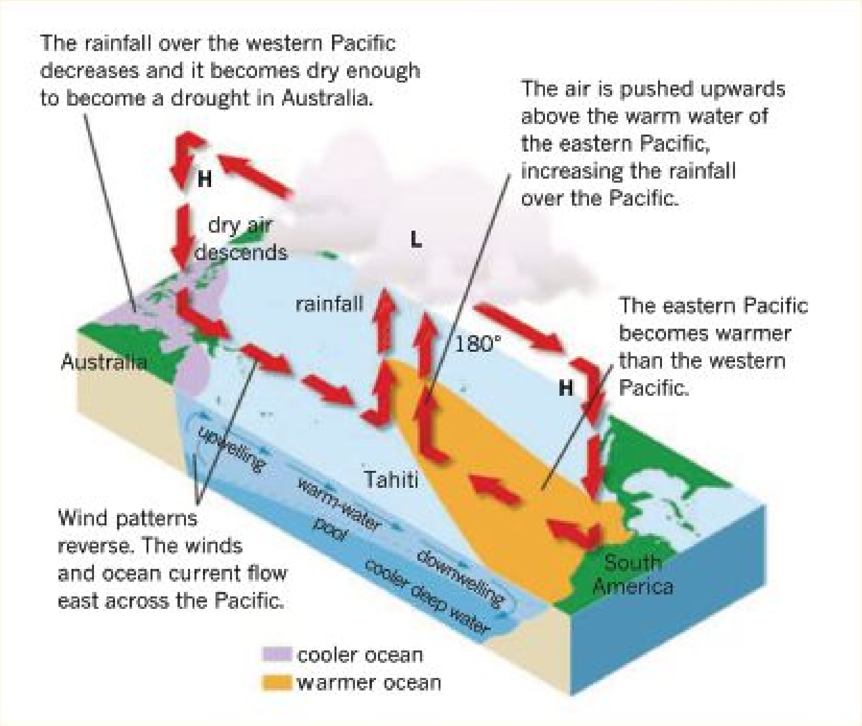


FIGURE 2.2.17 Conditions for an El Niño event. While South America may experience wetter conditions, large areas of Australia will experience hotter, drier conditions.

At the other extreme is a La Niña event, shown in Figure 2.2.18. During this event north-eastern Australia, Malaysia, the Philippines and Indonesia experience wetter conditions.

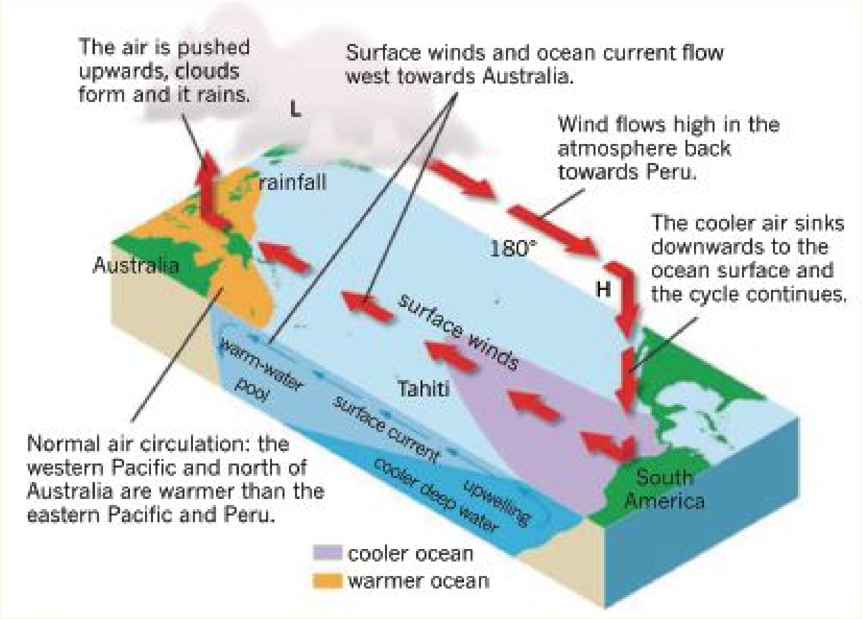


FIGURE 2.2.18 Conditions for a La Niña event. South America will be drier while eastern Australia will experience wetter conditions and stronger cyclones in northern regions.

Climate scientists are finding evidence that the enhanced greenhouse effect is causing a kind of double El Niño where the effects are more increased and last for longer periods. Normally, the warm water from an El Niño spreads across the Pacific and cools as it evaporates. The increased moisture in the air leads to thunderstorms and tropical storms. That hasn't happened as much as anticipated over recent years. The energy has not been taken out of the ocean, meaning that small temperature increases could hang around for a number of years. While those numbers may seem small, in the context of global climate, a shift of that magnitude could have a major effect.

Indian Ocean Dipole

The Indian Ocean Dipole (IOD) is a cycle of change in the water temperature between the eastern and western areas of the Indian Ocean that borders Australia's western coast. Its effects are not as strong as the Southern Oscillation. Cool surface waters in the Indian Ocean near Western Australia mean cooler, drier air. Hence there is less rainfall, particularly in central and southern Australia as shown in Figure 2.2.19(a).

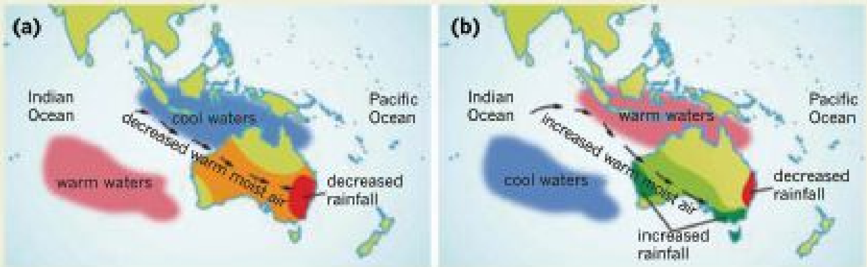


FIGURE 2.2.19 (a) Cooler water off the coast of Australia causes lower rainfall in central and southern Australia. (b) Warmer waters to the north cause increased rainfall.

When warm waters are near Australia, moist air circulates over north-western Australia, and southern regions can expect more rainfall. This can be seen in Figure 2.2.19(b). The size of the difference between the sea temperatures and how long this difference lasts will affect the length of dry periods.

2.2 Review

SUMMARY

- The overall temperature of the Earth is determined by the total thermal energy received, largely from the Sun, and the amount that is lost back to space, outside the Earth's atmosphere. Any change in that balance will lead to a warming or cooling of the Earth as a whole.
- Only about half of the incoming radiant energy from the Sun is eventually absorbed by the Earth's surface. This is because about 23% is absorbed by the atmosphere. Of the remainder, clouds reflect about 26% back towards outer space and the Earth's surface reflects about 4%.
- The energy that reaches the Earth's surface heats up the surface and is then partially radiated back out into space as longer-wavelength radiation.
- Specific gases present in the lower atmosphere in very small proportions absorb the long-wavelength radiation from the Earth and re-radiate it back to the Earth's surface.
- Evidence suggests that the atmosphere is absorbing and retaining more of the long-wavelength infrared radiation from the Earth's surface, in an enhanced greenhouse effect. This is due to increased concentrations of greenhouse gases.
- Evidence suggests the increase in concentration of greenhouse gases is due to human activities over the last 100 years.
- The rate at which the energy is re-radiated by the Earth will depend on the surface material.
- Large built-up areas can act as heat traps, increasing localised temperatures and retaining more thermal energy than rural areas.
- A small proportion of the thermal energy heating the Earth comes from the Earth itself.
- Movement of thermal energy through the Earth occurs by conduction at the surface through the crust and convection through the mantle deep within the Earth.
- Ocean currents are mass movers of thermal energy over very large distances. The large-scale ocean circulation of water is called the great ocean conveyor.
- Convection is the major process by which thermal energy is moved around the Earth's atmosphere.

KEY QUESTIONS

1 What is the main source of thermal energy heating the Earth?

2 Complete the paragraph below by choosing the correct response from the choices given in brackets.
Radiant energy from the Sun reaches the Earth largely as electromagnetic radiation in or near the [**infrared/visible/ultraviolet/radio**] wavelengths. The Earth reflects radiant energy as [**shorter/same/longer**] wavelength, [**infrared/visible/ultraviolet/radio**] radiation.

3 Why does carbon dioxide have such a big impact on the enhanced greenhouse effect?

4 Match the greenhouse gases to the sources of their production by humans.

Greenhouse gas	Source of greenhouse gas
carbon dioxide	air-conditioners
methane	artificial fertilisers
CFCs	agriculture
nitrous oxide	combustion of fossil fuels

5 Describe the major process by which thermal energy moves in the Earth's mantle.

6 Describe the major process by which thermal energy moves in the Earth's atmosphere.

7 Describe how sea breezes form.

2.3 Scientific modelling: The enhanced greenhouse effect

This section examines scientific modelling, uncertainties in data and the scientific process. These processes are applied to the ongoing modelling of the Earth's climate and the impact of the enhanced greenhouse effect. An example of just one type of data scientists collect concerns the salinity (salt content) of the oceans. The satellite image in Figure 2.3.1 uses colour coding to show different levels of salt concentration.

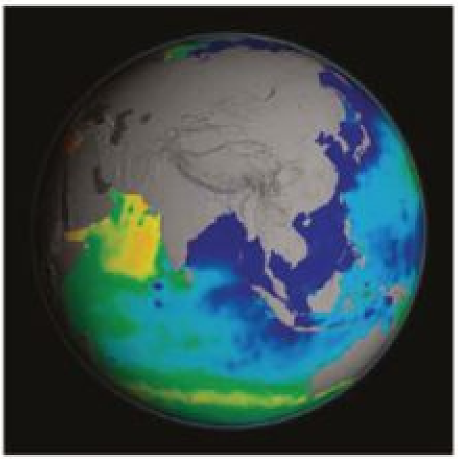


FIGURE 2.3.1 Aquarius satellite data for the salinity of the oceans around Asia in September 2014. The colours indicate high concentration (red) through to medium and low concentration (blue).

THE SCIENTIFIC METHOD

The **scientific method** is the process that scientists use to construct theories to explain practical observations. Following the scientific method means that personal and cultural influences that might affect findings are removed. The scientific method places importance on applying standard procedures, the critical review of other scientists' findings and the need to be able to replicate (repeat) previous tests with the same outcomes. These processes ensure there is no bias in data.

PHYSICSFILE

The beginnings of the scientific method

Galileo Galilei (1564–1642) was the first to develop and practise the scientific method. It is for this reason that Einstein referred to Galileo as the father of modern science. Galileo set up processes that allowed him to reliably compare measurements at different places and at different times. In other words, he conducted controlled experiments, which is the basis of the scientific method.

Galileo worked on concepts in a range of areas including physics, mathematics, engineering and astronomy. He observed four of the moons of Jupiter through one of the first optical telescopes. He noted that the moons were orbiting around Jupiter, suggesting that the Earth actually revolved around the Sun. This was a major break from the religious teachings of the time, which believed the Earth was at the centre of the universe. This separation of scientific theory from religion and philosophy was a major step in the development of scientific thought.

Key steps in the scientific method

The scientific method has four key steps:

1. observe and describe a particular phenomenon
2. form a **hypothesis** that explains the observations
3. apply the hypothesis to predict results of new tests
4. perform tests to assess the **reliability** of the predictions. This may lead to the acceptance of the hypothesis or to further investigation.

If the test results support the hypothesis then that hypothesis becomes regarded as a **scientific theory**. Otherwise the hypothesis is either modified to explain the new test results or abandoned. Theories that can't be tested are not accepted as scientific theory. Hypotheses for which the tests cannot be replicated (repeated) by another scientist are also regarded as not proven.

As scientists are continually gathering new data, no scientific theory can be thought of as fact. If repeated experiments consistently disagree with a current theory, then scientists will propose a new hypothesis that better fits the new observations. Scientists will always allow for the potential of a new hypothesis that better explains experimental results.

ANALYSING CLIMATE DATA: RELIABILITY AND UNCERTAINTY

Section 2.2 described the various means by which thermal energy moves around the Earth, i.e. predominately through the oceans and atmosphere. Together, the modes of transport and storage of thermal energy create an extremely complex system. The balance of the system can be completely changed by one seemingly unrelated factor.

Scientists must piece together seemingly unrelated pieces of information in order to refine a model that can accurately predict changes in the Earth's climate. Early models will always be uncertain. As more observations refine the model, the model becomes increasingly more accurate. However, the model is always liable to change in light of new observations and evidence.

See, for example, the model of Australia's climate shown in Figure 2.3.2. This model shows many different influences that can combine or contribute separately to Australia's climate.

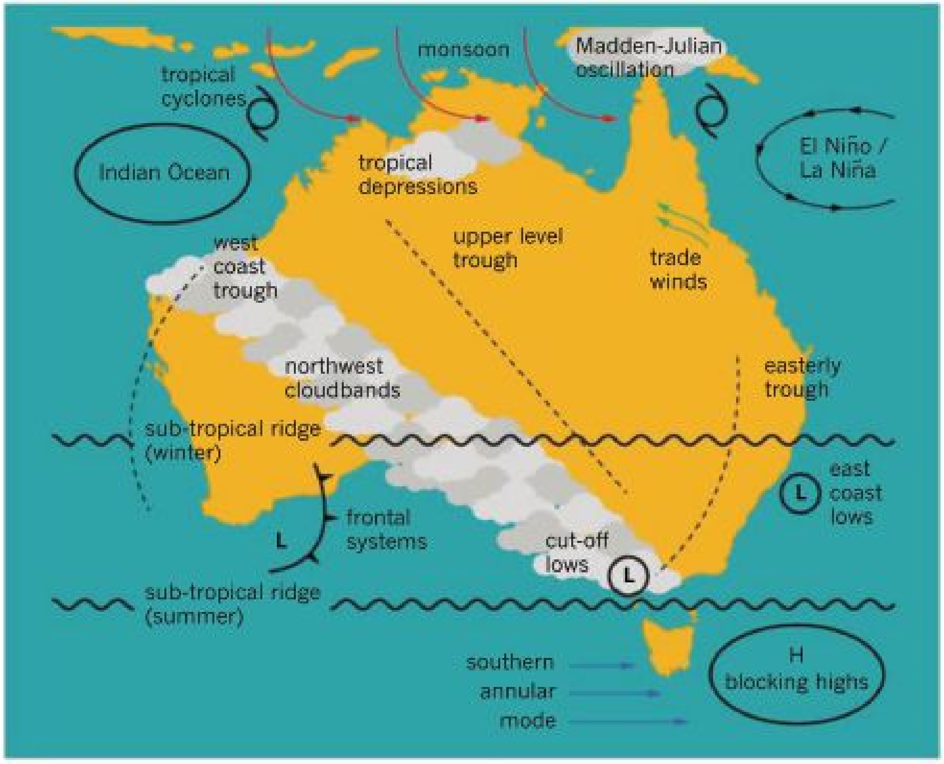


FIGURE 2.3.2 The main influences on Australia's climate. The combined effects can have different results in different areas of Australia at different times of year.

Formulating a definitive model to fully explain the effect of the enhanced greenhouse effect on climate change is very complex. This has led scientists to discuss many contributors to climate change.

Long-term climate trends

Scientists have had evidence for many years that the Earth's climate was once very different from what you experience today.

In order to model the Earth's climate in the past and make predictions about what can be expected in the future, scientists have employed a large number of different tools. Ice-core samples from Antarctica, tree rings, pollen levels, geographical features and long-term weather observations have been carefully analysed. Scientists use long-term data to identify long-term, naturally occurring variations in the Earth's climate. They compare this with new data to determine the significance of any changes in the new data they have identified.

PHYSICSFILE

Thawing permafrost

Many giant craters have been discovered in places such as Siberia in Russia. Scientists believe they result from thawing of the permafrost in recent, warmer-than-usual summers, which have been 5°C above the long-term average for that region. As the permafrost thaws, methane is released, building up gas pressure until it 'explodes' at the surface. The sound of running water and methane levels of 9.6% at the bottom of the craters provide practical evidence on which scientists are basing their initial theories.

FIGURE 2.3.3 An aerial photograph of a giant crater discovered in 2014 on the Yamal Peninsula in Western Siberia, Russia.

A recent event for which short-term data did not match long-term trends occurred on 24 March 2015. On this day it was considerably warmer in Antarctica than in Melbourne. A cold front cooled Melbourne while areas of Antarctica were experiencing temperatures more than 15°C above average. Short-term climate anomalies such as this are excluded from long-term modelling. A snapshot of the temperature on that day is shown in Figure 2.3.4.

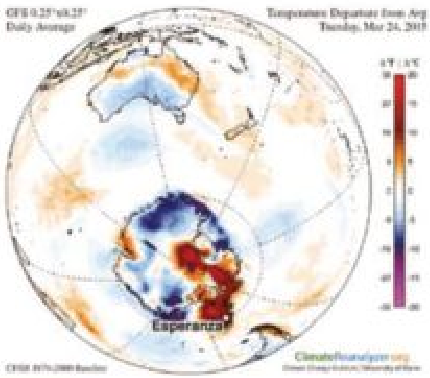


FIGURE 2.3.4 A meteorological image showing temperatures through Antarctica and Australia on 24 March 2015. The colour-enhanced image shows that temperatures in Antarctica were warmer than those experienced in Melbourne on the same day.

Warming and cooling

Scientists have found that there have been many cycles of warming and cooling in the Earth's history. In fact, 2.5 million years ago an ice sheet covered much of southern Australia, including Tasmania. Many cycles of warming and cooling followed. About 20 000 years ago a period of warming began, as shown by the graph in Figure 2.3.5. The graph also shows that the Earth is currently in a warming or **interglacial** period. Based on past records, interglacial periods last around 15 000 to 20 000 years before an ice age returns. If this long-term trend is to be used as a prediction, the Earth could be nearing the end of the current warming period.

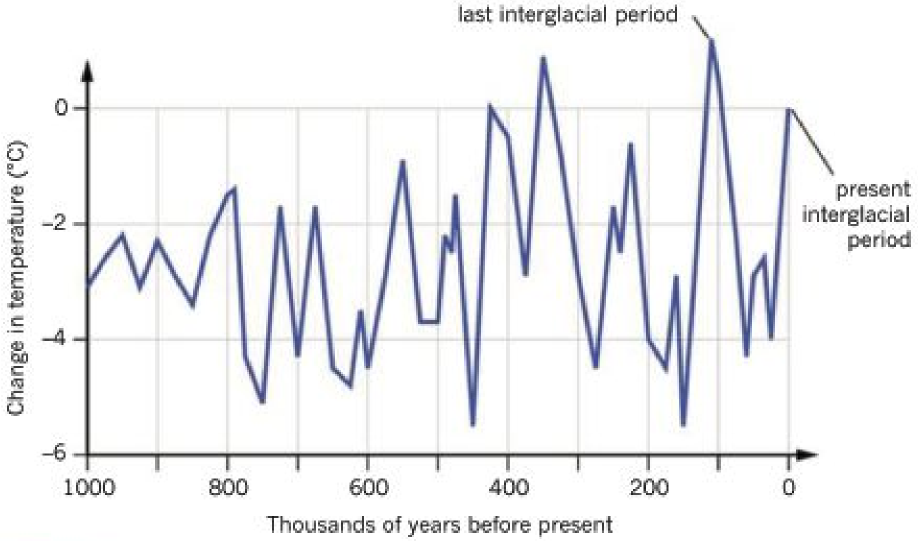


FIGURE 2.3.5 Long-term trends of warming and cooling. The change in temperature shown is based on the Earth's current average temperature.

The reasons for past global warming and cooling are not fully understood. This makes it difficult to completely understand exactly what effect humans are having on the natural processes of climate change. The current interglacial period started just before the start of the Stone Age, when the human population was small. Therefore, this current interglacial period cannot be explained by human actions alone.

Carbon dioxide levels

Carbon dioxide levels, a significant greenhouse contributor, have also varied considerably over Earth's history. Across geological history, the Earth has experienced times of significantly higher levels of carbon dioxide than current times. Scientists believe that 100 million years ago carbon dioxide levels were many times higher than now, although the exact value is uncertain.

In very general terms, long-term reconstructions of atmospheric carbon dioxide levels going back in time show that 500 million years ago, atmospheric carbon dioxide was likely around 20 times higher than current levels. (The long-term trend is displayed in the graph in Figure 2.3.6. on page 58) The levels fell, then rose again around 200 million years ago to 4–5 times present levels. It was in this period that there was a rise in giant fern forests. This event led to the laying down of the carbon deposits that have become Earth's current sources of fossil fuels. It was also a period of higher temperatures. Carbon dioxide levels then started a slow decline until the Industrial Revolution.

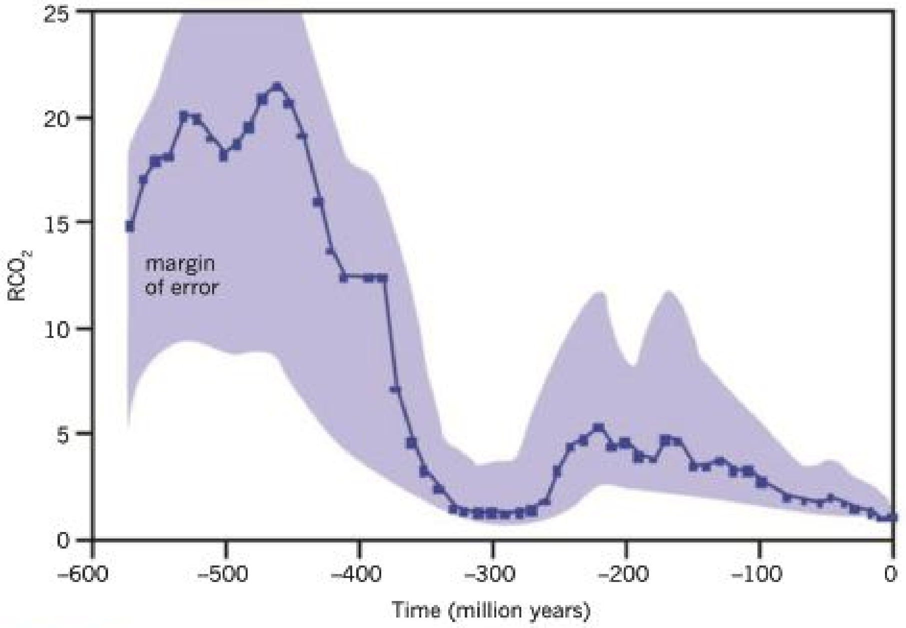


FIGURE 2.3.6 The history of atmospheric carbon dioxide through geological time. The parameter RCO$_2$ is defined as the ratio of the mass of carbon dioxide in the atmosphere at some time in the past compared to that at present (with a pre-industrial value of 280 parts per million).

Another analysis of carbon dioxide levels, shown in Figure 2.3.7, looks only at more recent history: the time span when our human ancestors were evolving into the modern human. Analysis of air bubbles trapped in ice cores in Antarctica allows an accurate measure of the proportions of carbon dioxide in the atmosphere over the past 400 000 years. During ice ages, carbon dioxide levels were around 200 parts per million (ppm). Over the warmer interglacial periods, they hovered around 280 ppm.

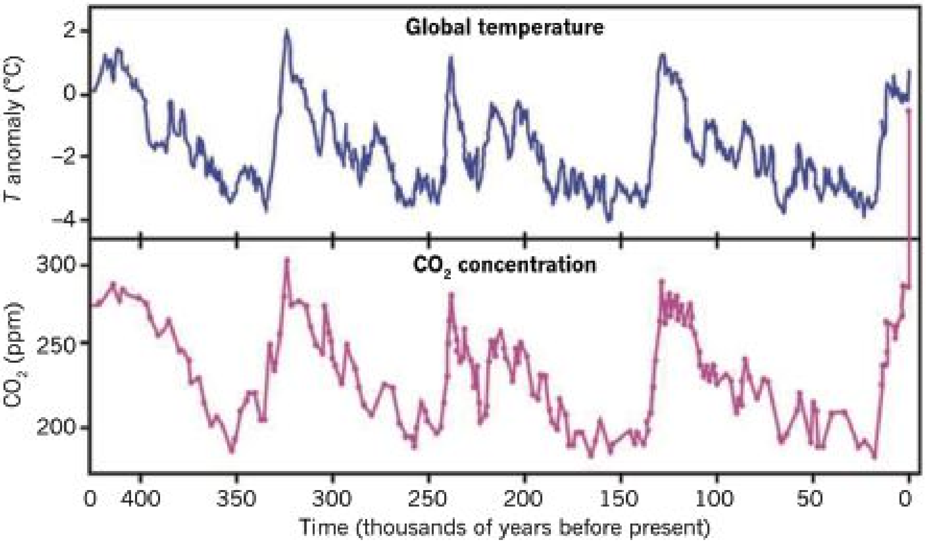


FIGURE 2.3.7 A comparison between carbon dioxide levels and global temperatures over the last 400 000 years.

Comparing changes in carbon dioxide with changing temperature shows a connection between higher carbon dioxide levels and the higher temperatures of interglacial periods. What can't be said is whether the increases in carbon dioxide caused the higher temperatures or whether they were a response to change. However, the recent increase in carbon dioxide levels since the beginning of the industrial age shows a very strong correlation (connection) to fossil-fuel burning. In 2014, carbon dioxide levels passed 400 ppm for the first time in human evolution.

Recent climate trends

It is very easy to explain trends based on limited data and to assume the explanations are correct. That approach differs from the scientific method. While scientists may form a hypothesis based on a limited set of data, the hypothesis must be repeatedly tested against new data, tests and observations.

In recent years, scientists have been gathering vastly greater sets of data on which climate models can be tested and refined to give more certainty about the human influence on an enhanced greenhouse effect.

Carbon dioxide levels have been monitored on a daily basis at the summit of the volcano Mauna Loa, Hawaii, since 1958. The data is presented in graph form in Figure 2.3.8.

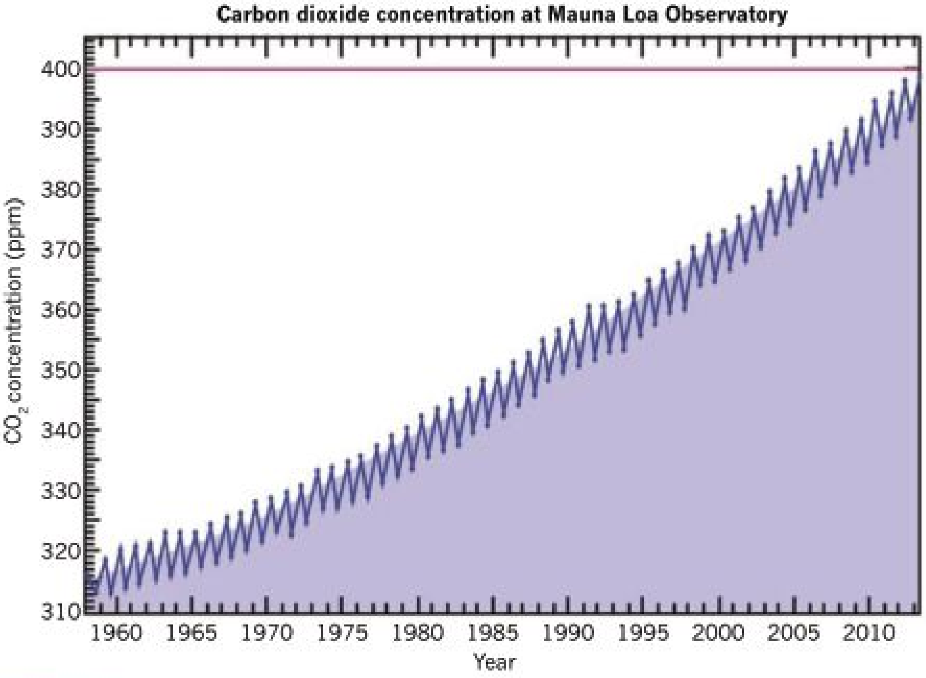


FIGURE 2.3.8 Atmospheric concentration observations of carbon dioxide at Mauna Loa Observatory since 1958.

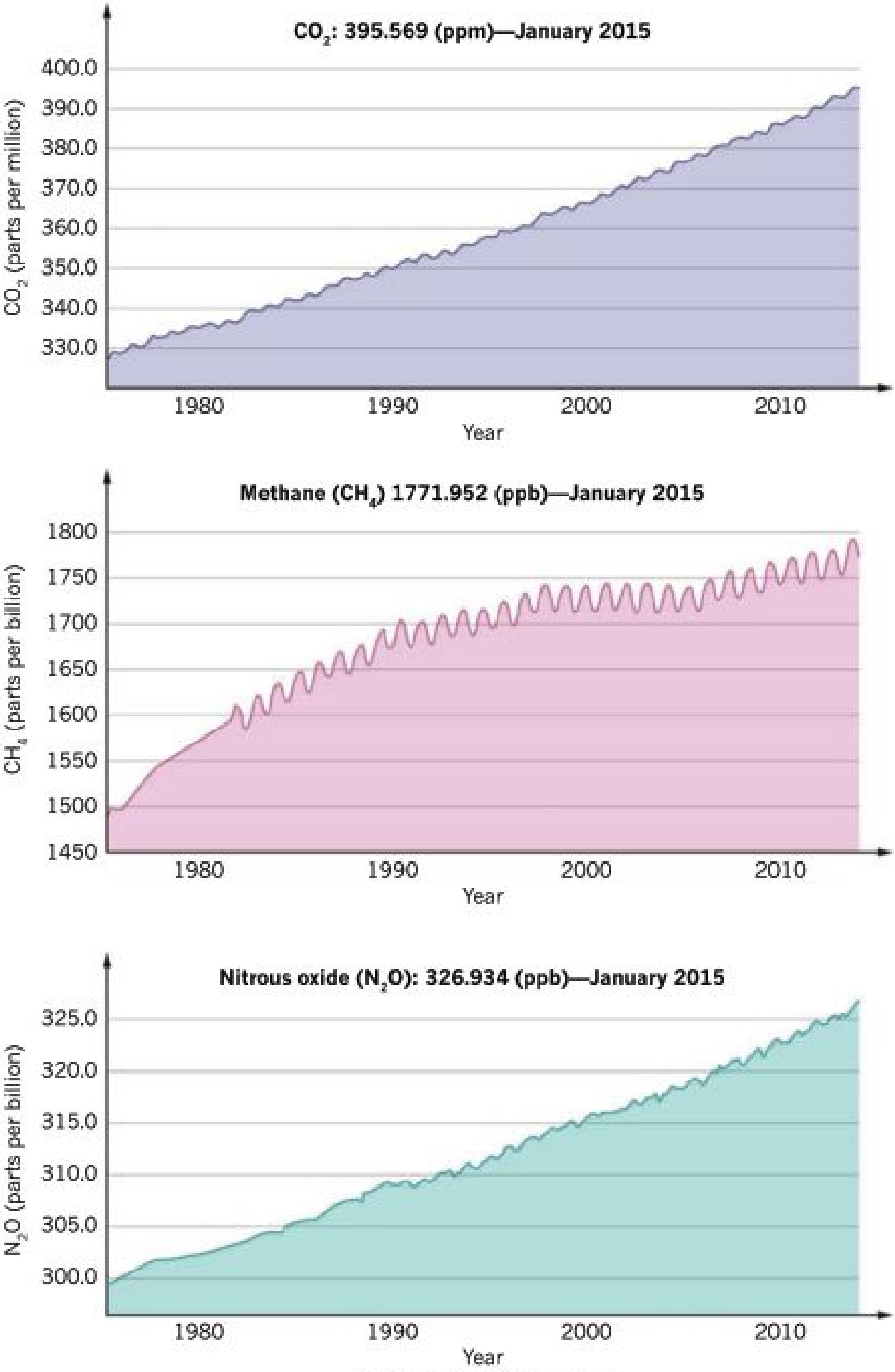
Mauna Loa Observatory (MLO) is an atmospheric research facility. It was established in the early 1950s in a remote location that takes advantage of the undisturbed air and minimal influences of human activity. The reason for the saw-tooth pattern in Figure 2.3.8 is that extra carbon dioxide is absorbed by the deciduous trees of the northern hemisphere in spring as they grow new leaves. This leads to small changes over the year.

Mauna Loa Observatory records carry particular weight because of the remote location, the length of time data has been collected and the consistency with which daily records have been kept. Scientists have used the same peer-reviewed methods to record data every single day since 1958. When recordings began at MLO, carbon dioxide levels averaged 315 ppm, an increase of around 10% from pre-industrial levels. Now, the average has reached 400 ppm. The last time levels were around this same level, humans did not exist, temperatures were an average of 3°C warmer and sea levels were at least 5 metres higher.

Mauna Loa Observatory is one of three premier baseline air pollution stations in the World Meteorological Organisation–Global Atmospheric Watch network. A second base station is located in Australia at Cape Grim on the north-west tip of Tasmania. The air sampled there is some of the cleanest in the world, having travelled long distances over the open Southern Ocean and being untainted by pollution from nearby cities or industry. Established in the early 1970s, data collection has not been running as long as Mauna Loa but reveals similar trends in atmospheric concentrations of greenhouse gases.

As the graphs in Figure 2.3.9 show, Cape Grim data confirms the trends from the Mauna Loa data.

- Greenhouse gases show annual cycles.
- Since 1976, carbon dioxide levels have increased by more than 15%. Concentrations of methane and nitrous oxide have increased by about 20% and 8%, respectively.
- Measurements from baseline pollution stations combined with ice-core samples from Antarctica reveal that carbon dioxide levels have increased by around 40%, methane levels have increased by more than 150% and nitrous oxide levels have increased by about 20% since the Industrial Revolution.



Last updated: March 2015

FIGURE 2.3.9 Air sample data collected at Cape Grim, Tasmania, under clean-air (baseline) conditions and analysed by CSIRO.

Looking for evidence of enhanced climate change

Earth's climate has always been changeable. There have been seven ice ages in the last 650 000 years. These have occurred as the Earth's climate has responded to changes in radiant thermal energy due to very small changes in the Earth's orbit. This point is raised by people who do not agree with the theory that human activity is causing the current warming trends. However, based on past cycles we are overdue for a cooling period, and the amount of warming over the last 150 years is unprecedented since human civilisation began.

Scientists have known about the heat-trapping ability of greenhouse gases since the mid-nineteenth century. Data collected from baseline stations and ice cores from Greenland, the Antarctic and tropical mountain glaciers all show a significant increase in the concentration of key greenhouse gases that clearly coincides with the start of the industrial age. Scientists also know from observations of ice cores that the Earth responds to changes in the concentrations of greenhouse gases very quickly—in human rather than geological time scales.

However, it is less easy to work out whether individual changes relate to natural climate change or to a human-caused enhanced greenhouse effect. The Earth's climate is incredibly complex. For scientists to be confident that a change is caused by human activity rather than natural variation, observations must support the hypothesis.

Any one change could have apparently unexpected or contradictory effects, as the following examples explain.

Ocean temperatures

The Earth's oceans have absorbed much of the additional thermal energy retained as a result of increased greenhouse gas concentrations. It is estimated that sea surface temperatures have increased by an average of 0.2°C since 1969. That might not sound like a lot, but given the high specific heat capacity of water it represents a very large amount of thermal energy.

At the same time a marine science expedition found that the deep waters of the Southern Ocean were cooler and less salty than 10 years before. This would seem to be contradictory.

One possible answer is the melting of the polar ice caps. Scientists know that Antarctica lost around 152 km^3 of ice between 2002 and 2005. As ice melted the resulting water would be both cooler and less salty than the surrounding oceans. This would explain the changes observed in Southern Ocean temperatures.

Whether through the influence of human activity or natural variation, such a change could affect the great ocean conveyor belt that influences regional and global climates. Any changes would influence a large region of ocean in just a few years by potentially blocking the conveyor belt from carrying warm water from the Caribbean, along the North American east coast and on to Northern Europe. North America did report a longer winter in early 2015, which would support this hypothesis.

Combined evidence

Modelling of global surface temperatures shows that:

- the Earth has warmed since the advent of the industrial age
- the majority of the warming has occurred since the 1970s
- the warmest 20 years have all occurred since 1981
- all ten of the warmest years have occurred in the past 12 years through to 2015, with 2014 the warmest year yet recorded as a global average.

The graph in Figure 2.3.10 shows the variation in global average temperatures from the long-term average. The graph also includes a trend line that clearly shows an increase in temperature from the long-term average.

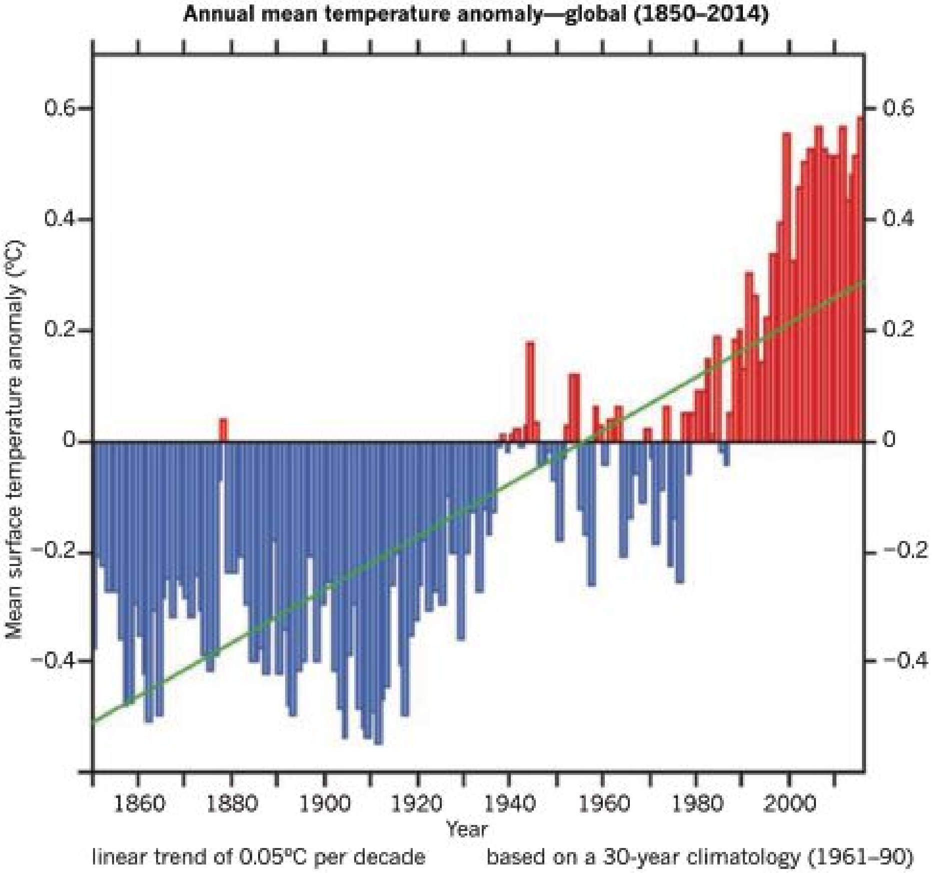


FIGURE 2.3.10 The annual mean temperature anomaly globally, 1850–2014.

There has been considerable debate over the **validity** of the measures of the Earth's surface temperature. For example, there are claims that the stations used for recording temperatures are in areas that have been affected by localised warming, such as urban heat traps. However, the three methods used to derive the data are largely independent, as the scientific method requires.

The conclusion of the Intergovernmental Panel on Climate Change (IPCC), that global warming is undeniable, does not rest solely upon records of increasing surface temperature. Increased surface temperatures are only one line of evidence among many, including:

- the uptake of heat by the oceans
- the melting of glaciers
- the rise in sea levels
- increased atmospheric surface humidity.

If the land surface temperature records were flawed and the globe had not really warmed, then it would be almost impossible to explain these changes.

One of the issues that has been raised in the media is the view that global warming stopped in 1998. This view ignores the presence of standard variation in the climate that scientists have long acknowledged. The El Niño event of 1998 was particularly strong—probably one of the strongest of the twentieth century. It was followed by cooling between 2006 and 2008, likely driven by a La Niña effect that occurred at that time. The area of cooler sea surface temperatures that defines La Niña conditions can push global temperatures down when the event is strong enough. This was coupled with a very low period of solar activity, also usually associated with decreased temperature. Despite this, every year of the twenty-first century so far has been warmer than the 1990 average, and as of 2014 the last 38 years in a row have been above the twentieth century average.

2.3 Review

SUMMARY

- The scientific method has four key steps:
  1 observing a particular phenomenon
  2 forming a hypothesis
  3 applying the hypothesis to predict other results
  4 testing the reliability of the predictions.
- In modelling climate data, scientists apply the scientific method to that data.
- Scientists piece together seemingly unrelated pieces of information in order to refine a model that can accurately predict changes in the Earth's climate. Early models will be uncertain. As more observations refine the model, the model becomes increasingly more accurate, but is always open to change in light of new observations.
- Scientists use long-term data to identify long-term naturally occurring variations in the Earth's climate and to determine the significance of new data that deviates from long-term trends.
- The Earth's climate has always been variable. Reasons for past warming and cooling are not fully understood.
- The recent increase in carbon dioxide since the beginning of the industrial age shows a very strong correlation with fossil-fuel burning.
- In 2013, carbon dioxide levels passed 400 ppm in the northern hemisphere for the first time in human evolution. This is well outside the variations in long-term data that we would expect.

KEY QUESTIONS

1 Which of the following is the correct definition of a scientific theory?

A It is an explanation for a phenomenon that has been tested and supported.

B It is the observation of a particular phenomenon that has not been tested.

C It is the proposed explanation of a phenomenon that cannot be tested.

D It is a proposed explanation for a phenomenon in which there is some uncertainty.

2 In order to find the percentage increase or decrease in methane concentration, values were taken from April 1997 and again in April 2014. Why is it important that values were used from the same month in each year?

3 Why is it important for scientists to follow proper scientific method in their investigations?

4 List the correct order of the steps for the scientific method.

5 The following text explains the relationship between carbon dioxide levels and long-term trends of warming and cooling. Complete the paragraph below by choosing the correct response from the choices given in brackets.

There is good correlation between increased concentrations of [**water vapour/carbon dioxide**] and [**warmer/cooler**] temperatures over the last 450000 years. During [**ice ages/interglacial periods**], carbon dioxide levels were around 200 parts per million (ppm). Over the warmer [**ice ages/interglacial periods**] they hovered at 280 ppm or more.

2.4 Issues related to thermodynamics

Chapters 1 and 2 of this text so far have concentrated on developing an understanding of heat and heat-transfer methods. By applying these thermodynamic principles to real-life contexts, it is possible to explain changes in the thermal energy of systems. An important way scientists use this knowledge is in describing, explaining and addressing the enhanced greenhouse effect. They also use this knowledge to bring about more sustainable ways to live. An example of using thermodynamic principles to improve sustainability is by harnessing the Sun's energy using panels like those in Figure 2.4.1.

FIGURE 2.4.1 Solar panels use the energy from the Sun to generate electricity.

PASSIVE SOLAR HOUSING DESIGN

Rising energy costs and dwindling resources have led to a re-evaluation of the use of non-renewable fuels. While the focus is often on industrial efficiencies, a lot can be achieved at the domestic level (at home). One way that Australians have reassessed their environmental impact is through designing homes using passive design. **Passive design** is energy efficient and requires little or no mechanical heating or cooling. Heating and cooling would normally account for about 38% of the energy use in an average Australian home. Homes that are designed with passive solar heating and cooling in mind take advantage of the natural climate to maintain thermal comfort. The principles of good passive solar design include a number of elements.

Insulation and thermal mass

Recall from Chapter 1 that thermal energy is always transferred from a hotter body to a cooler body. So, in winter, the aim of good passive design is to limit the transfer of thermal energy from a house (heat loss). In summer, the aim is to limit the transfer of thermal energy to a house (heat gain).

The use of insulating materials in a passive design home prevents heat loss by conduction in winter and heat gain by conduction and radiation in summer.

This therefore provides a more stable temperature, which requires less-active forms of heating and cooling. There are two main ways to maintain stable temperatures in the home:

- reduce air leakage—air leakage accounts for 15–25% of winter heat loss in buildings and can contribute to significant loss of cool air in hot climates. Sealing a home will ensure less mechanical heating and cooling is used, reducing energy bills and greenhouse gas emissions.
- increasing thermal mass—floors and walls that absorb and store thermal energy are particularly useful for naturally heating and cooling homes. Thermal mass refers to materials that retain or store the energy provided by sunlight. Large amounts of energy are needed to change the temperature of high-density materials such as concrete, bricks and tiles. These materials take longer to heat up on hot days and to cool down on cold nights.

Orientation

Orientation refers to the way in which a house is positioned on its site to take advantage of climatic features such as sunlight and cooling breezes. This can play a large role in how much heat is lost or gained. For example:

- positioning large windows on the northern side of a house allows generous amounts of solar radiation to warm the interior in winter
- small windows to the south provide ventilation but reduce heat loss on the shady side of the house
- positioning living areas on the northern side allows the winter sun to warm areas of the house where it's needed during the day
- bedrooms that are little used during the day can be sited on the southern side of the house.

The diagram in Figure 2.4.2 shows the ideal placement of windows to aid passive design.

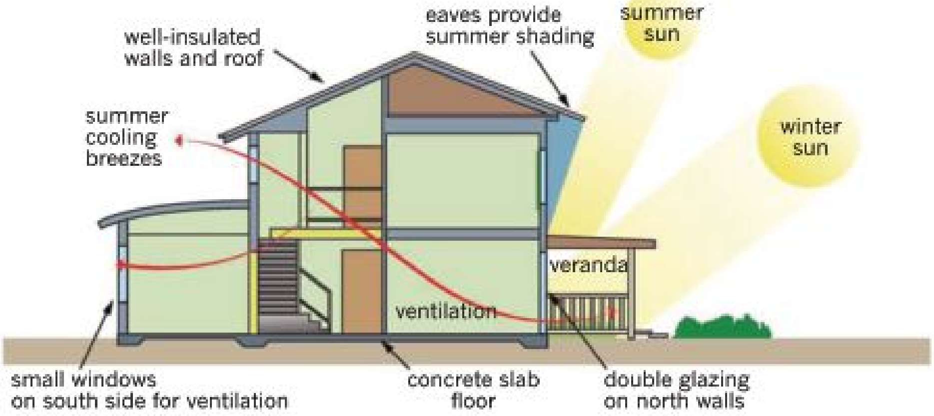


FIGURE 2.4.2 The passive design home.

Shading

Windows in a typical insulated home can account for more heat gain or loss than any other element in the building. In summer, heat gain through an unshaded window can be 100 times greater than through the same area of insulated wall. One square metre of ordinary glass can let in as much thermal energy as would be produced by a single bar radiator. In winter, heat lost through a window can be ten times more than through the same area of insulated wall. Effective shading can include:

- overhanging eaves
- window awnings
- shutters
- pergolas
- deciduous trees.

For example, blank walls heavily shaded by trees or eaves block the sun on hot summer afternoons and shield against the cold winds of winter.

Eaves are the parts of the roof that overhang the exterior walls, as shown in Figure 2.4.2. They play an important role in controlling the amount of sunlight that reaches the walls of a house. Eaves block more sunlight in summer than in winter. This is because the sun is higher in the sky during the summer. In winter, when the sun is lower in the sky, sunlight is able to reach a greater portion of the wall.

Calculating sun angles for the specific geographic location can also increase the effectiveness of these methods to maximise the shade effects in summer and avoid blocking the winter sun.

Glazing

Large windows provide light, ventilation and daytime heating. At night or on cold, cloudy days windows are a major source of heat loss. Glazing (glass) has a major impact on the energy efficiency of the whole building. On a winter's night when it is 15°C colder outside, an average home (with 70 m^2 of clear glass windows, glazed doors and aluminium frames) will lose thermal energy at a rate of approximately 6.5 kW. That's equivalent to the total heat output of a large gas heater or an air conditioner running at full capacity. The majority of that heat loss is by conduction through the glass of the window. For this reason, windows are a major part of energy-efficient home design.

Insulating glass, usually in the form of double glazing, helps to stop the transfer of heat by conduction. Double glazing provides resistance to this type of heat transfer by utilising a sealed space between the panes of glass, as shown in Figure 2.4.3. The air in the gap between the glass panes is a poor conductor of heat, which means that much less thermal energy will be transferred through a window that is double glazed.

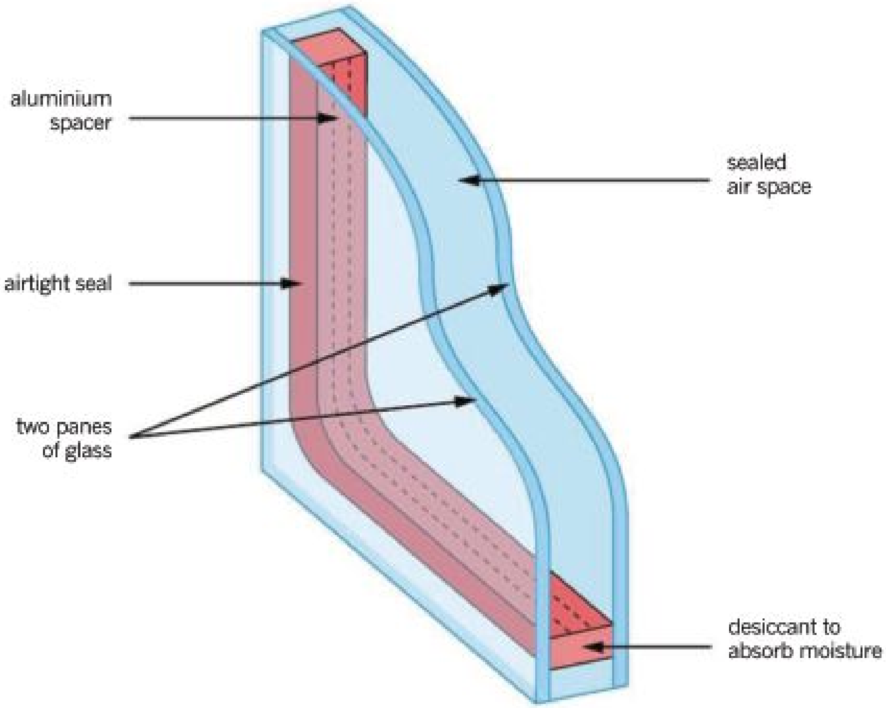


FIGURE 2.4.3 The air space between the panes of glass in double-glazed windows limits thermal conduction.

Increasingly, argon gas is being used to fill the space between the glass panes instead of air. As well as having a lower conductivity than air, argon is also widely available and cheap. Good use of double glazing can reduce heat loss or gain by 50%, or more when additional treatments are added to reduce thermal transfer by radiation.

PHYSICS IN ACTION

The Sustainability Learning Centre, Hobart

The Sustainability Learning Centre in Hobart, shown in Figure 2.4.4, has been designed to put the principles of passive solar design and low **carbon footprint** into practice. Its use of renewable and natural resources like sunlight, rainwater and cooking oil make it a very energy-efficient building

FIGURE 2.4.4 The Sustainability Learning Centre in Hobart has been designed using the principles of passive solar design and has a low carbon footprint.

Low carbon footprint

The centre was constructed to have a low carbon footprint. Some of its design features include:

- a solid floor construction that does not use concrete, and a roof system with minimal structure
- materials that involve minimal processing were used where possible
- minimal use of concrete, steel and aluminium. The production and transport of these materials has a high carbon footprint.
- extensive use of timber. The use of timber in construction acts as a carbon sink by preserving timber and the carbon stored in it, as the original tree converted it via photosynthesis from carbon dioxide in the atmosphere.

Recycled materials have been used extensively in the centre. The use of recycled materials significantly lowers the amount of energy, including fossil fuels, used in the processing of new materials. The Sustainability Learning Centre undertook recycling in the following ways.

- Many of the materials for the Centre were salvaged from an old warehouse in Hobart. Much of this material would otherwise have become landfill. Instead, 98% of the materials were saved, and the bulk has been used in the centre—including 80% of its steel and timber, and half of its concrete.
- Most of the Centre's timber framing, brickwork, roofing, external cladding, paving and joinery are recycled.
- Much of the furniture in the centre has been recycled from other buildings.

Overall, the use of recycled metals and timbers has saved 15 tonnes and 4 tonnes of carbon dioxide emissions, respectively.

Energy efficiency

The design of the Sustainability Learning Centre achieves a 111% energy saving when compared to a benchmark in the same climate zone. The way lighting is used contributes to part of this saving.

Lights in the Centre are controlled by motion and light-level sensors that automatically turn the lights off when they are not required. Lights near windows operate at a lower intensity than those further away from natural lighting.

The Centre is designed to use the sun's energy to heat the buildings as much as possible.

- The entire north-facing side of the building acts as a **solar collector** with a **trombe wall**, a **perforated solar collector wall** and double-glazed windows.
- Large panels of glass let in the sun, which heats the dark stone floors and bricks, which then release that heat throughout the day.
- High levels of insulation and thermal mass trap and store the sun's energy for heating.

Photovoltaic cells and vacuum tube solar collectors generate electricity and heat water for use by the occupants. The building is designed to produce more energy to put back into the grid than it takes from the grid, making it a net producer of energy.

The Centre makes use of natural daylight and ventilation to further reduce energy consumption. When passive heating isn't enough, a **hydronic heating** system is used. The circulating water in the hydronic heating system is heated by a furnace that operates on recycled cooking oil. On hot days, the building is cooled by the night sky. Rainwater is pumped up onto the roof at night where it naturally cools. The cooled water is then pumped down and through the floor pipes to cool the tiles under foot.

DOMESTIC HEATING AND COOLING SYSTEMS

Domestic heating and cooling systems use different principles of thermodynamics in order to operate effectively. Each type, however, has an impact on the environment and contributes to the enhanced greenhouse effect.

Comparing the operation and efficiencies of different systems can help us to make informed choices and to understand the impacts of human activity. Some of the systems currently used are:

- heat pumps, also known as reverse-cycle air conditioners
- resistive heaters
- evaporative coolers
- solar hot-water systems
- electrical resistive hot-water systems.

Heating

As a first option for heating a home, it is best to incorporate some passive design principles. This will vastly decrease the need for other forms of heating, which have large impacts on energy consumption and therefore on the environment.

When active forms of heating are required, the two main types of heaters are radiant and convection. Radiant heaters largely heat people and objects by direct radiant energy. The air itself absorbs little of the energy. Convection heaters warm and circulate the air in the room. Other forms of heating, such as hydronic floor heating, also heat by conduction via direct contact.

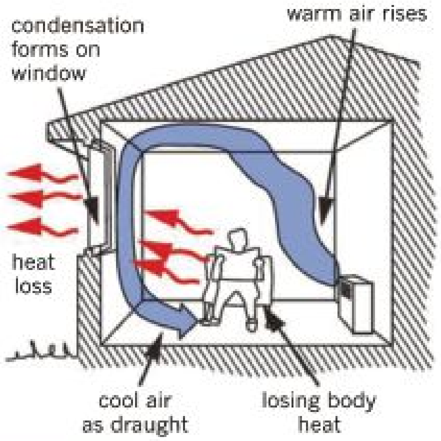


FIGURE 2.4.5 Movement of air in a home is created by heaters as hot air rises and then falls as it cools. Hot air loses heat through conduction when it comes into contact with uninsulated surfaces.

All heaters produce air movement as hot air naturally rises. Air is cooled when it comes in contact with windows and poorly insulated walls and ceilings. The cool air then falls. This cycle is shown in Figure 2.4.5. To minimise this cooling effect, it is important that walls, windows and ceilings are well insulated. You can also stay warm by dressing appropriately. Up to 5% of heating bills can be saved by dressing more warmly and by reducing the thermostat setting on a heater by just 1°C.

Different forms of heating are best for different circumstances.

- In large rooms with high ceilings, a combination of radiant and convective heating is best.
- In small rooms, convective heating is the most efficient.
- In large, draughty rooms and bathrooms, radiant heating is the best option.

Heat pumps or reverse cycle air conditioners

Reverse-cycle air conditioners are commonly known as heat pumps. They work by transferring heat from one place to another.

Inside a heat pump, a volatile liquid known as a refrigerant is pumped through a closed circuit of pipes. The refrigerant is called volatile because it very easily changes state. A typical configuration of these pipes is shown in Figure 2.4.6.

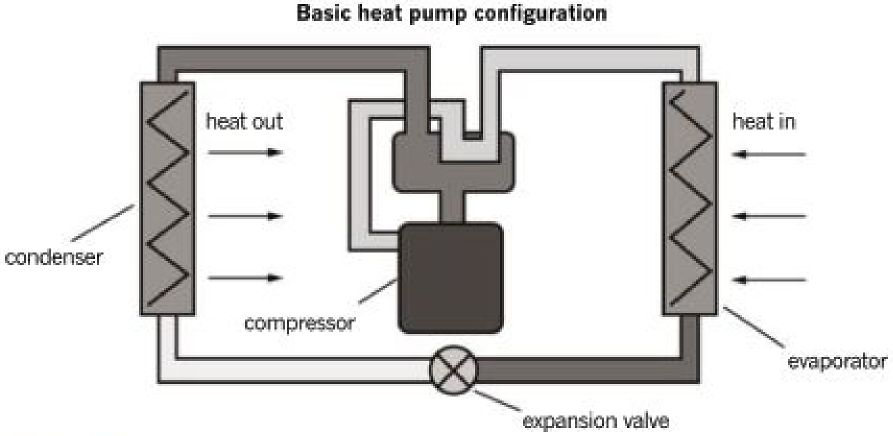


FIGURE 2.4.6 A reverse-cycle air conditioner works as a heat pump. Reversing the process will heat the inside of a house. Heat pumps can also be used in water heaters.

The basic operation of a heat pump is as follows:

- The liquid refrigerant passes through an expansion valve, which lowers the pressure of the refrigerant so that it turns into a gas.
- As it expands, the refrigerant cools down. Note that the amount of thermal energy that exists in the refrigerant has not changed but the temperature has dropped because the thermal energy is less concentrated.
- Remember that thermal energy always moves from higher temperatures to lower temperatures. Therefore it is important that the temperature of the refrigerant is significantly lower than the temperature of the air that is to be cooled down.
- A fan draws air from the hot room over these very cold coils.
- The air cools down due to conduction, and it is then blown back in the room.
- The refrigerant now has the thermal energy that it removed from the air.
- The refrigerant then passes through a compressor that squeezes the gas, putting it under increased pressure. This raises the temperature of the refrigerant because the thermal energy becomes more concentrated. The purpose of raising the temperature of the refrigerant is to force it to give up thermal energy in the next step.
- The refrigerant then passes into the condenser pipes that are exposed to the outside air. Because the compressor raises the temperature of the refrigerant, the refrigerant is warmer than the outside air and so thermal energy is transferred by conduction from the refrigerant to the outside air.
- By transferring thermal energy to the outside air, the refrigerant cools down and condenses to a liquid state.
- The cycle then repeats.

A heat pump is an energy-efficient means of heating as it relies on small energy inputs, around 20% on average, in addition to that provided by the local environment. The environment itself is supplying the bulk of the heat. Only a small pump is required to circulate the refrigerant and to change the pressure. The heating cycle is generally the most efficient mode for heat pumps to work.

Gas heaters

Gas heaters can either be permanently installed or portable. They basically work by burning natural gas, a fossil fuel. Ducted gas heaters pass cold air over a heat exchanger, which is warmed by gas combustion. The warmed air is then pushed through ducts into the home through a series of vents in the ceiling or floor where it heats the room through convection.

A portable gas heater works in the same way, but pushes heat out as radiant heat. In ducted systems, the products of combustion are released outside the home; in a portable gas heater they are released into the room, which can be very dangerous.

Even though natural gas is a fossil fuel, it is the cleanest-burning fossil fuel. It creates less greenhouse gas emissions than by heating oil or generating electricity from fossil fuels.

Electric heaters

Small electric heaters can be useful when heating just one or two rooms. These heaters come in a variety of types. They can be cheap to buy but are often very expensive to run. Some of these heaters are not very effective and their inefficiency can result in high greenhouse-gas emissions, unless they are combined with a renewable energy source. The main types of heaters include radiant bar heaters, fan heaters, convective heaters and oil-filled column heaters.

- Electric radiant heaters convert electrical energy into radiant heat energy though the use of heating elements that reach a very high temperature. The element is usually inside a reflector that directs the energy away from the heater.
- Electric fan heaters work in the same way to provide instantaneous heat. They combine an electric element with a fan to blow warm air around a room.

- Oil-filled column heaters use electricity to heat the oil that is sealed inside columns. This heat is then transferred to the steel casing that circulates the warmth through natural convection (without a fan). These heaters also emit some radiant heat.

Most of Australia's electricity is derived from burning coal. This generates huge amounts of carbon dioxide, which is released into the atmosphere. Unless powered by renewable energy, these small electric heaters can produce up to six times the greenhouse gas emissions of an efficient gas heater.

In-slab central heating

In-slab or underfloor heating systems have either an electrical element or hot-water pipes built into a concrete floor, which effectively turns the floor into a giant radiator. They provide a combination of radiant, convective and conductive heating. The floors are slow to warm and cool due to the high thermal mass of the concrete slabs. Therefore they are poorly suited for buildings or rooms where heating is only needed occasionally, and in changeable weather where more heat is needed quickly.

Wood fires

In rural areas, wood can be an excellent energy source because it is renewable if sustainably harvested. However, in urban areas wood fires contribute to poor air quality. Additional emissions are associated with the transport of firewood to areas where firewood can't be sustainably grown and harvested.

Energy choices for heating

Gas heaters and reverse-cycle air conditioners (or heat pumps) produce only one-third of the greenhouse gas emissions of standard electric heaters. The most efficient 5- and 6-star reverse-cycle units produce one-fifth of the emissions of conventional electric heaters. They are the most efficient and environmentally friendly forms of active heating.

Table 2.4.1 compares the cost and greenhouse gas emissions for different heating systems.

System type	Running costs	Greenhouse gas emissions
high-efficiency ducted natural gas	low	low
hydronic zoned natural gas or heat pump	low	low
ducted reverse-cycle or heat pump	medium	medium (low when using renewable energy sources)
hydronic zoned with wood/solar heat source	low	very low (based on sustainable wood harvesting or solar)
in-slab high off-peak electric	medium–high	very high (low when using renewable energy sources)
electric heaters	high	very high (low when using renewable energy sources)

TABLE 2.4.1 Cost and emission comparisons for various heating systems.

Cooling

The better insulated, shaded and draught-proofed a house is, the smaller the cooling unit needed and the less often it will need to run. Shading windows with fast-growing deciduous trees is one way that the cooling of existing homes can be improved. However, if a home has been poorly designed for cooling or has little space around it for planting suitable trees, mechanical cooling is more likely to be needed.

The three major methods of mechanical cooling are:

- fans
- evaporative coolers
- air conditioners.

Fans

In a well-insulated home with good ventilation, fans alone can often supply enough cooling. Fans don't actually reduce the temperature or humidity in the air. Instead, they increase air flow, which increases convective cooling. Moving air around the body creates a similar effect to reducing the actual air temperature by around 3°C (refer to 'Wind chill' on page 23 in Chapter 1). Fans are also the cheapest units to buy and have the lowest greenhouse-gas emissions.

Evaporative cooling

Evaporative air conditioning uses the process of evaporation to cool the air. A pump circulates water from the reservoir on to a cooling pad, which becomes very wet. A fan draws hot air from outside the unit through the moistened pad. As it passes through the pad heat is transferred from the air to evaporate the water. Therefore, the resulting air is cooler. Evaporative air conditioners use a very small amount of electricity to power the pump. As a result they are a very environmentally friendly cooling option.

However, evaporative cooling isn't effective in all conditions. For the cooling effect to be noticeable, the humidity outside needs to be relatively low. In Victoria, the humidity on hot days can often be above the level at which an evaporative cooler works efficiently.

Air conditioners

As mentioned previously, modern air-conditioning systems are also called heat pumps. In the case of cooling, thermal energy is transferred from the space to be cooled to the outside air.

These systems are usually called reverse-cycle air conditioners because they can be used to transfer heat into the room from the outside, or reversed to transfer heat out of the room. This is a particularly effective and energy-efficient method of heating and cooling.

Geothermal heat exchangers

While heat pumps run on air-to-air heat exchange, a far more efficient method is air-to-water or air-to-ground heat exchange. This kind of heat transfer occurs in geothermal heat exchangers. This is not a mechanical system; it relies on a difference in temperature to move thermal energy in and out of a house. It works by running heat-exchange pipes into a stored water supply or deep underground, where the temperature is relatively constant throughout the year. Unlike conventional heat pumps, they continue to work efficiently in extreme weather conditions and can also be used to run hot-water services.

As can be seen in Figure 2.4.7, the geothermal heat exchanger keeps the indoor temperature steady all year round.

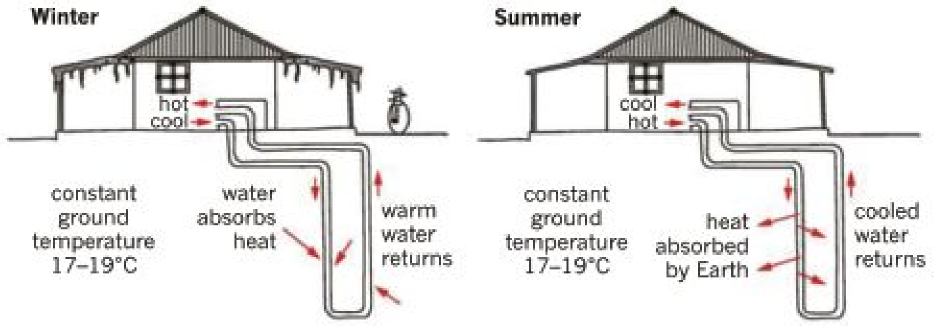


FIGURE 2.4.7 Geothermal cooling and heating systems are very efficient heat exchangers and retain that efficiency even in extreme temperature events.

Solar air cooling systems

Solar air cooling systems use a fan or ventilator to extract hot air out of the roof space between the reflective foil under the roof and the metal roof sheets. They work by extracting hot air from the roof space and replacing it with ambient (surrounding) air. Reducing the temperature inside the roof minimises heat transfer to the ceiling space below. The effectiveness of removing hot air from the roof space is, however, very sensitive to roof colour and the presence of the reflective foil.

Table 2.4.2 compares the cost and greenhouse gas emissions for different cooling systems.

System type	Running costs	Greenhouse gas emissions
fan	low	low
evaporative cooler	low	low
refrigerated cooler (heat pump)	medium	medium
air conditioner (non-heat pump)	high	high
geothermal	low–medium	low–medium
solar air cooler (with electric fan)	low	low

TABLE 2.4.2 Cost and greenhouse emission comparisons for various cooling systems.

Water heating

Hot water accounts for around 16% of the average Victorian household's energy costs. Therefore an energy-efficient water heater has the potential to reduce household energy usage and environmental impact significantly.

Solar water heating

An average household roof of around 200 m^2 receives 3600 MJ of energy in 5 hours of direct sunlight. Given that an average household might use around 72 MJ of electrical energy in a day, it is possible to power the whole house from sunlight alone.

A simple way to make use of this free source of energy is to harness the energy from the Sun and use it to heat water. All solar water systems use the same basic principle. A solar collector panel is placed on the sunniest side of the roof and is connected to a water storage tank. The solar collector is usually made up of dark-coloured pipes that allow the flow of a fluid. In warmer climates, such as Australia's, the fluid is regular drinking water, which becomes the hot water used within the home. In cooler climates, the fluid is often a mix of water and antifreeze, which stops the water freezing in the pipes. This water–antifreeze mixture is pumped through a water storage tank in closed pipes, where it transfers the thermal energy to the water for use in the home.

There are two main types of solar water heaters.

- Active solar water heaters that use electrical pumps to move the hot water around the system.
- Passive systems that rely on convection currents to move the hot water. The natural convection currents used in passive systems are a low-energy means of moving hot water from a solar collector to a storage tank without the use of an electric pump. Gravity acts to draw cold water from the bottom of the storage tank down through a collector panel that heats the water. By the time the water reaches the end of the collector panel it is warmer and therefore less dense, so it rises back to the top of the tank for re-circulation. A simple passive system is shown in Figure 2.4.8.

FIGURE 2.4.8 A passive solar water heater. Gravity draws cold water through a solar collector panel, and the less-dense hot water then rises back to the top.

Electric hot-water systems

In a standard electric storage hot-water system, a heating element in the bottom of the tank heats the water for storage in an insulated tank. When the water at the bottom is heated, it becomes less dense than the surrounding cold water. The warm water rises and convection currents form in the tank. In time, all of the water in the tank is heated and is then available for use during the day. Generally, the water is heated at night using lower-cost off-peak electricity. The water is heated to a temperature of around 60–70°C, which is higher than is needed for most household applications. When hot water is used, it's drawn from the top of the tank and the tank is refilled with cold water from the bottom. A thermostat senses the drop in temperature and the heating element is turned on to reheat the water.

As with electric heaters, the majority of the energy used to heat the water comes from the burning of coal. This has a large impact on the enhanced greenhouse effect through the release of carbon dioxide.

Gas hot-water systems

Storage tanks can also be heated by natural gas. Gas storage heaters use a gas burner located under the storage tank to heat the water. They work in the same way as electric hot-water systems, but natural gas is a cleaner-burning fuel and therefore has a lower impact on the environment.

Heat-pump systems

A heat-pump water system works by transferring heat in the air outside to water stored inside a storage tank. It works in exactly the same way as a heat-pump system is used to heat a house. A fan draws outside air into an evaporator, which has pipes inside that are filled with a refrigerant. The warm air vaporises the refrigerant, which is then compressed, raising its pressure and temperature.

A heat exchanger transfers the heat from the heated refrigerant gas pipes to a tank where water is stored. As the refrigerant vapour condenses back to its liquid form, it heats the stored water. Once the refrigerant is back to a liquid state, the heat-pump cycle begins again.

This is a very efficient way to heat water and has a low impact on carbon dioxide levels and the enhanced greenhouse effect.

FUELLING A CAR

In the very early days of motoring, the internal combustion engine using petrol as a fuel was not the only means of propulsion being considered. Steam cars were common until German engineer Karl Benz combined the internal combustion engine and some clever mechanics to basically define the car as it is today. Only during the world wars did engineers consider alternatives to petrol because of fuel shortages. There are many reasons why environmentally friendly and sustainable fuels are gathering popularity, including:

- the environmental impact of burning fossil fuels which create greenhouse gases
- the increasing cost of oil
- the need for Australia to import fuel.

Petrol engines have a high environmental impact because they are powered by non-renewable resources. They also produce some harmful emissions. Among other things, petrol engines release carbon monoxide, oxides of nitrogen and unburnt hydrocarbons into the atmosphere. All of these products negatively impact on the environment.

Table 2.4.3 lists some of the common fuels used in vehicles today and the amount of carbon dioxide produced by burning each fuel.

Fuel type	kg of CO_2 per unit
petrol	2.38 kg per litre
diesel	2.65 kg per litre
LPG	1.67 kg per litre
ethanol	1.5 kg per litre
grid electricity	43 kg per kW h

TABLE 2.4.3 Carbon dioxide emissions from common fuels.

The use of different fuels is not as clear cut as it may first appear. It is important to note that the carbon-emission figures quoted in Table 2.4.3 are purely the emissions created by the use of the fuel in the vehicle. These figures do not allow for:

- the energy used in fuel extraction
- the energy used in fuel refining
- the construction of cars and their maintenance
- the construction of roads and their maintenance.

A study in the United Kingdom suggests that the true value for carbon emissions should be scaled up by around 32% based on an average driving distance per car per year (14,000 kilometres). That would give a carbon emission value of 3.15 kg per litre for petrol and of 3.48 kg per litre for diesel.

A range of alternative fuels are now in production or under development.

Diesel

While standard petrol-driven passenger cars remain the most popular, the sale of diesel cars has increased in recent years. As a result, Australians are now using more than 1 billion litres less petrol per year than in the early 2000s. Diesel outsold petrol in 2014 by more than 15%.

Diesel-powered vehicles have quite a considerable advantage over petrol-powered cars in terms of economy. The typical family diesel car uses around 20% less fuel per kilometre than a comparably powered petrol car.

However, in terms of environmental impact the picture is much less clear cut.

- Diesel produces 2.7 kg of CO_2 per litre, whereas petrol produces 2.4 kg of CO_2 per litre.
- Burning diesel produces significantly greater proportions of nitrous oxide, another greenhouse gas and a contributor to acid-rain production. Burning diesel also produces other **particulates** (pollution) that are potentially harmful to humans and the environment.

While many manufacturers of diesel cars have now moved to filter out particulates, choosing diesel over petrol is not a clear-cut choice.

LPG

Liquefied petroleum gas (LPG) is the generic name for mixtures of gaseous hydrocarbons (mainly propane and butane). LPG is synthesised (derived or produced) from petroleum or natural gas and, as such, it comes from a non-renewable fossil-fuel source.

LPG has many benefits, such as:

- it burns more cleanly than petrol and diesel
- it produces fewer particulates than diesel
- it produces hardly any sulfur emissions.

Although LPG is cleaner than petrol:

- it has less energy and so is less efficient as a fuel
- it is an invisible and highly flammable gas, making careful storage and delivery essential
- synthesising the gas puts pressure on non-renewable resources.

LPG conversion is usually done to petrol cars already on the road, and it can be done to many petrol cars. Some manufacturers have also produced cars that have been factory fitted to use LPG, usually as a duel-fuel option to standard petrol.

Ethanol

Ethanol is an alcohol produced by fermenting a range of crops such as sugars and grains. In Brazil, ethanol blended with petrol has been a popular fuel for many years. In recent years it has also gained popularity in Australia. In Australia, ethanol is usually mixed with standard petrol for use in cars and trucks. E10, or fuel containing 10% ethanol, has been available for some years. More recently, cars have been released in Australia that are able to run on E85, which is 85% ethanol.

Potentially an E85 blend provides a more sustainable and renewable resource: 1 litre of ethanol produces 1.5 kg of CO_2 when burnt as a fuel. This compares very favourably with petrol and diesel which produce 2.4 kg and 2.7 kg of CO_2, respectively.

However, there are still downsides to using ethanol as a fuel.

- Increased demand for ethanol as a fuel has driven up food prices.
- Ethanol production has led to further land clearing to grow more of the crops needed for fermentation. Land clearing itself is a major contributor to carbon emissions and an enhanced greenhouse effect.
- Ethanol as a fuel has 33% less energy per litre than petrol. To produce the same amount of energy, 1.4 litres of ethanol is required instead of each 1 litre of petrol that would have been used. Calculating the overall amount of CO_2 produced by ethanol for the volume of petrol required for the same amount of energy suggests that the overall difference in CO_2 produced is small.

However, all of this ignores the extra benefit of consuming a plant-based fuel. Taking into account all of the production and distribution factors, it is estimated that ethanol-blended fuels save around 70% of CO_2 production compared with straight petrol.

Ethanol-blended fuels are usually cheaper and have a higher octane rating than standard petrol—95 compared with 92 for standard petrol. A higher octane rating means that the fuel can be used in higher compression engines, resulting in greater fuel efficiency. However, as the compression rating of the engine is usually set for standard petrol, using ethanol doesn't always give greater power and efficiency. Most cars manufactured since 1986 can use ethanol-blended fuels, but only those manufactured specifically for blended fuels are likely to benefit.

Biofuel

Ethanol is a biofuel, but the term biofuel is more closely associated with fuels that can be used with standard diesel engines. Therefore biofuel is also referred to as biodiesel. Biodiesel is refined from vegetable oil, animal fats, waste cooking oil and other sources of hydrocarbons.

Biodiesel has many advantages.

- It is sustainable.
- It has a negligible or a very low effect on greenhouse-gas emissions. This is because it is recycling carbon rather than releasing trapped carbon from fossil fuels.
- It is cleaner burning than standard diesel, producing lower emissions and fewer particulates.
- Being a diesel fuel, it also offers better fuel economy than petrol.
- It is biodegradable.
- It is non-toxic.

The big drawback currently is the cost of production. Biodiesel will not become a truly viable fuel until better, cheaper production methods are developed.

Most of the major vehicle manufacturers are exploring ways to produce biodiesel in greater volumes at a competitive price. If biodiesel could be produced at close to the price of diesel and with minimal environmental impact, it would mean that many of the cars currently in use could use it without expensive modifications to existing diesel engines. This would result in significantly lower total emissions.

One of the alternative methods of biodiesel production being explored is using algae to make ethanol and biodiesel, as shown in the image in Figure 2.4.9. A benefit of producing oil from algae in this way, rather from plant oil crops such as rape seed oil, is that it does not take up space that could be used for growing food crops and it does not use food crops that could be feeding people. The overall environmental impact on land clearing and on food production is considerably reduced. This method is still in development and is not yet able to produce fuel in sufficiently large volumes, but it does hold promise for a more sustainable future.

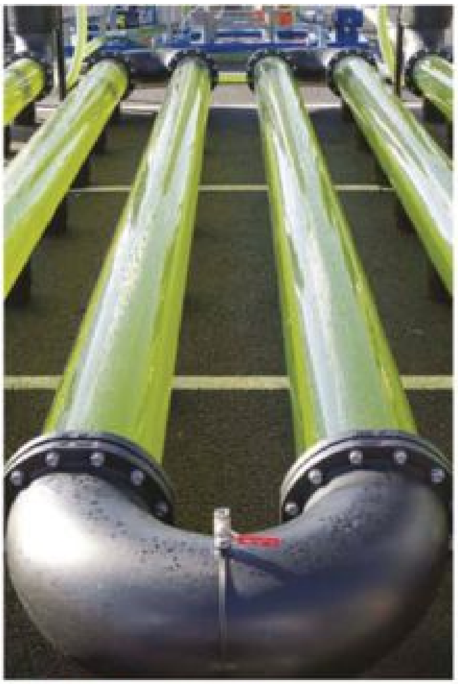


FIGURE 2.4.9 An AlgaeLink algae-growing system being used to make ethanol and biodiesel.

Electricity

The term 'electric car' usually refers to a vehicle that runs on an electric motor. The motors take power from a large battery that needs to be plugged in to a mains power supply for recharging. Solar power and hydrogen fuel cells can also be used to generate electricity to power a car as a hybrid technology, in combination with a separately fuelled engine.

The advantages of using electricity are considerable:

- running costs are low
- electric motors run quietly
- electric motors create zero direct exhaust emissions.

However, using currently available technology, the drawbacks are also considerable:

- electric vehicles are expensive
- electric vehicles have a limited range, constrained largely by existing battery technology
- recharging causes a potentially large environmental impact. In Victoria, plugging into a non-renewable electricity source using coal-fired power stations creates large amounts of greenhouse gases. The greenhouse-gas emissions are comparable with the direct use of fossil fuels.

Electric cars will only reach their full environmental benefit when they are used in conjunction with sustainably produced electricity supplies.

There are not many pure electric cars, such as the Tesla (Figure 2.4.10) or the Nissan Leaf, on the road today; hybrid vehicles are more common. A hybrid system uses a combination of a battery-powered electric motor and a petrol or diesel combustion engine.

FIGURE 2.4.10 The Tesla Model S electric sports car. The Tesla is the first fully electric car that offers both high performance and a reasonable range of up to 600 km between charges.

Hydrogen

A hydrogen fuel cell creates electrical energy by combining hydrogen and oxygen. This process creates a fully sustainable, environmentally neutral system. The only waste material produced is water vapour.

Hydrogen fuel-cell technology could lead to fuel-efficient and emission-free cars. Small model cars have been running on fuel cells for many years, and many car manufacturers have passenger cars under development or in limited production. But there are still many problems to overcome before fuel cells are a viable alternative.

There is no hydrogen-fuelling infrastructure yet available in Australia to allow passenger cars to refuel (see Figure 2.4.11). In Perth, busses powered by hydrogen fuel cells were introduced in 2004. They were withdrawn again in 2007 because the technology didn't suit Australian conditions. The higher cost meant that the economics of running the buses couldn't be justified.

FIGURE 2.4.11 A liquid-hydrogen filling station in Germany. The test car is being filled with liquid hydrogen fuel (from the large white tank on the right) at a hydrogen filling station. The converted car can run for 300 km on a full tank of liquid hydrogen.

Reducing the environmental costs of fuelling a car

One consistent way of reducing running and environmental costs, regardless of the fuel type, is to reduce the size of the engine itself. By reducing the size of the engine, the total mass of the car is reduced. Less pulling power is needed and the engine itself becomes more economical to run. Table 2.4.4 on page 78 shows the typical carbon-dioxide production for common-sized passenger vehicles, along with other forms of transport.

Fuel type	kg of CO_2 produced per litre of fuel
small petrol car 1.4 litre engine	0.17 per km
medium car 1.4–2.0 litre engine	0.22 per km
large car	0.27 per km
average petrol car	0.20 per km
small diesel car (<2.0 litre engine)	0.12 per km
large diesel car	0.14 per km
average diesel car	0.12 per km
articulated truck, diesel engine	2.68 per km
rail	0.06 per person per km
air, regional	0.18 per person per km
air, long haul	0.11 per person per km
shipping	0.01 per tonne of freight per km

TABLE 2.4.4 The carbon-dioxide production of some common vehicles.

Public transport vehicles are substantially more efficient than cars, even cars with small engines. While a large passenger train uses vastly more fuel in total than any passenger vehicle, the overall amount of carbon dioxide produced per passenger is about a third that of a standard petrol-powered passenger car.

ENERGY USE IN THE HOME

National energy use

While new homes must be built to energy-efficient standards, very few Australian homes are yet to be totally heated and cooled by passive solar sources. Figure 2.4.12 shows the energy usage of an average household in Australia. As you can see, 38% of that energy is spent on heating and cooling (or HVAC as it is called: heating, ventilation, air conditioning).

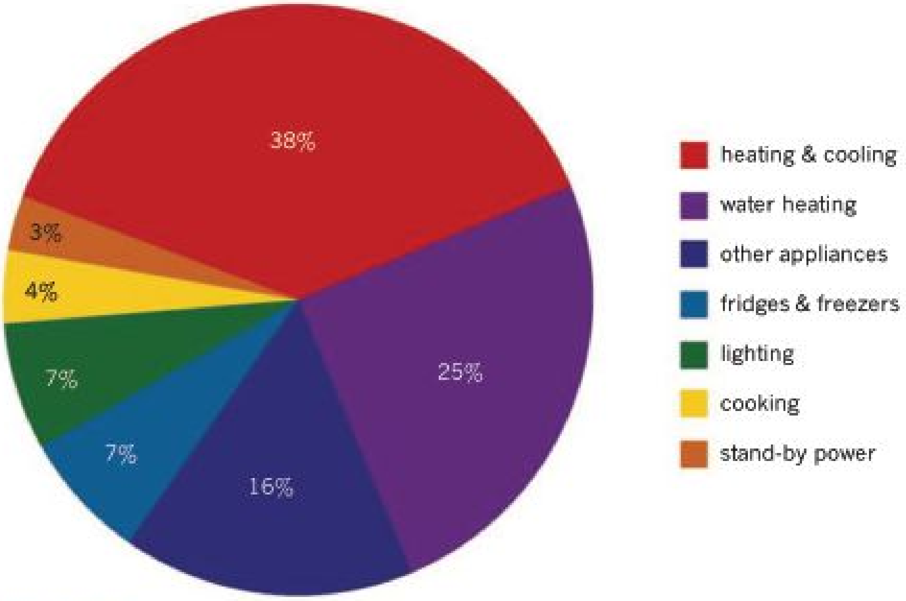


FIGURE 2.4.12 The proportions of energy use by category for the average Australian home. The proportion Victorians are using on heating and cooling is well above the national average and is increasing.

PHYSICSFILE

Star ratings

Energy ratings have helped to improve the energy efficiency of many standard household appliances. Refrigerators, dishwashers and other appliances are now all sold with a clearly indicated energy efficiency rating based on a star system. The more stars, the more energy-efficient the appliance is compared with other similar appliances. However, as households have become more affluent, the size of appliances has increased, offsetting much of the gains. For example, in 1966 the average household refrigerator had a capacity of 245 litres and only 15% of households had more than one. In 2007 the average capacity of a household refrigerator had increased to 366 litres and almost one-third of households had two.

In Victoria, heating and cooling accounts for around 60% of annual energy use, well above the national average. The peak demand for electricity in Victoria has been growing much faster than the overall electricity demand, which has generally declined as power costs rise and a greater proportion of homes install solar panels. The main reason for the increased energy use is the use of air conditioners.

While energy use for heating has remained fairly static, energy used for cooling homes is increasing rapidly. This is because more homes are installing air conditioners to manage the increasing numbers of extremely hot days over summer. Air conditioners run exclusively on electricity and add to problems of meeting peak demand, particularly in cities. While peak demand doesn't stay at those levels for long, it is peak demand that drives the construction of new power stations and power distribution networks. This means increased costs and huge environmental impacts. The emissions from the use of air conditioners contribute to the levels of greenhouse gases in the atmosphere, which in turn drives temperatures upwards.

Figure 2.4.13 shows just how much extra energy Australians are using than in the past. This is predicted to rise to 467 petajoules (1 petajoule = 10^{15} J) by 2020.

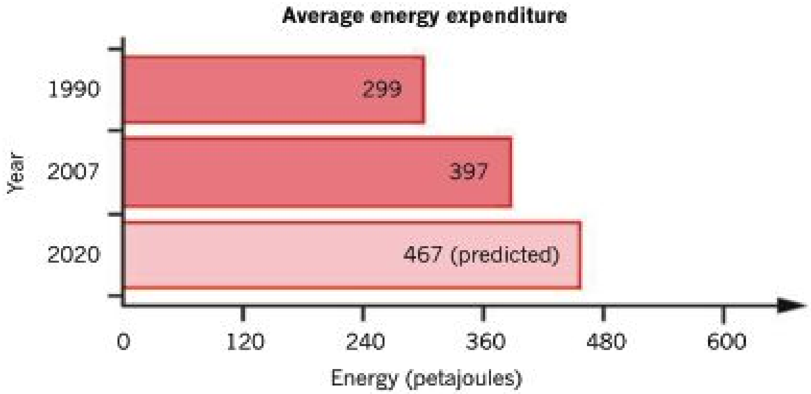


FIGURE 2.4.13 The average energy use across Australia continues to rise. This is mainly attributed to household cooling.

PHYSICSFILE

Other energy use in the home

Air conditioners get much of the blame for the growth in household energy consumption, but the use by other appliances is also growing. Televisions were once a negligible proportion of household energy use. Recently there has been an increase in the number of households with one or more large-screen televisions. This has resulted in a sharp increase in the amount of energy used by televisions. It is projected that by 2020, large-screen televisions will account for 25% of the total energy used by household appliances. The total energy used by all appliances is expected to reach 169.3 petajoules, at which the level of energy use by appliances will be similar to that used for home heating.

2.4 Review

SUMMARY

- Passive solar design uses thermodynamic principles to reduce the energy consumption of homes.
- Passive solar design includes:
  - increasing insulation and thermal mass
  - paying attention to the orientation of the building on the site and the positioning of rooms within the house
  - increasing shading around the building
  - using double glazing and other window treatments.
- Domestic heating and cooling systems use different principles of thermodynamics to operate effectively.
- Heating systems include:
  - heat pumps
  - gas heaters
  - electric heaters
  - in-slab heating
  - wood fires.
- Cooling systems include:
  - fans
  - evaporative cooling
  - air conditioning
  - geothermal heat exchange
  - solar cooling systems.
- Different heating and cooling methods have different benefits and disadvantages based on running costs, energy efficiency and greenhouse-gas emissions.
- Vehicle fuel choices, such as diesel, petrol, LPG, ethanol, biofuel, electricity and hydrogen, can affect the greenhouse-gas emissions from cars.
- The efficient use of fuels in vehicles can also result in lower greenhouse-gas emissions.

KEY QUESTIONS

1 Which design feature of a home can account for more thermal energy loss or gain than any other?

2 For good passive heating of homes in Australia, homes should be sited with large windows on the northern side of the house. Why is the north side preferred in Australia?

3 Which of the following is a feature in passive design?

A insulation and thermal mass
B orientation
C shading
D glazing
E all of the above

4 Through which heat-transfer method(s) does hydronic in-slab heating heat a room?

5 Sort the following list of fuel alternatives from the smallest producer of carbon dioxide per litre to the highest:

diesel, LPG, ethanol, petrol

6 How much of the energy used in the average Australian home is used on heating and cooling?

Chapter review

02

KEY TERMS

black body
carbon footprint
electromagnetic radiation
emissivity
enhanced greenhouse effect
fossil fuels
geothermal energy
greenhouse effect
greenhouse gases
hydronic heating
hypothesis
interglacial
isthmus
mantle
Pangaea
particulate
passive design
perforated solar collector wall
photovoltaic cell
radiation
reliability
scientific method
scientific theory
solar collector
trombe wall
validity

1 All objects with a temperature above absolute zero emit radiant thermal energy. As an object's temperature increases, what happens to the wavelength and frequency of the emitted radiation?

2 List the seven categories of electromagnetic radiation from the highest energy to the lowest.

3 A light globe is labelled as 'cool daylight' and another as 'warm white'. The 'warm white' globe appears more yellow than the 'cool daylight'. Based on this observation, which of the statements below is correct?

- **A** The apparent temperature of the 'warm white' globe is higher than the one marked 'cool daylight'.
- **B** The apparent temperature of the 'warm white' globe is lower than the one marked 'cool daylight'.
- **C** Both globes have the same apparent temperature.
- **D** The apparent temperature of the globes can't be determined from the information available.

4 The power to a radiator is turned down and the surface temperature of the heater element halves as a result. What effect will this have on the rate of transmission of radiant thermal energy (power) from the radiator?

5 A black solar water heater has an emissivity of 0.98. What does this imply about the heater's ability to

- **a** absorb radiant thermal energy?
- **b** emit radiant thermal energy?

6 The centre of the Earth is estimated to be at a temperature of 7000 K. Using Wien's law, determine the peak wavelength of the radiant energy emitted at this temperature.

7 Two black kettles are at the same temperature, approximately 50 K above the surrounding temperature. They have the same colour and surface characteristics. The first kettle has double the surface area of the second. At what rate will the first kettle emit radiant energy compared with the second?

8 The emissivity, *e*, of a surface depends on the surface colour, the surface characteristics (matte or shiny, for example) and the temperature of the surface. Which of the following statements about the value of *e* for a particular surface is true?

- **A** A matte black surface will have a value of *e* close to 0.
- **B** A shiny, white surface will have a value of *e* close to 1.
- **C** A shiny, white surface will have a value of *e* close to 0.
- **D** None of the above.

9 The floor of the Sustainability Learning Centre in Tasmania is heated by the Sun's radiation to a comfortable 30°C. The floor re-emits this thermal energy as radiation, heating the room. What would be the peak wavelength of the re-emitted radiation?

10 The quartz element of particular radiant heater glows orange at a peak wavelength of 650 nm. Based on this wavelength, what temperature, in kelvin, is the element of the heater?

11 Ashleigh was conducting an experiment to determine the emissivity of a particular metal surface. She found that when the surface was heated to 1000 K, the rate of emission of thermal radiation was 45 360 W. If the metal surface was 1 m^2, what was the emissivity of the metal surface?

12 The surface of a particular star is at a temperature of 9300 K. With what region of the electromagnetic spectrum does the peak wavelength of the radiant thermal energy for this temperature coincide?

13 What does the enhanced greenhouse effect refer to?

14 How do high-density urban areas and land clearing contribute to the greenhouse effect?

Chapter review *continued*

15 Select the correct statement about the total radiant thermal energy re-radiated by the Earth.

- **A** about 5% is directly lost to space
- **B** about 50% is directly lost to space
- **C** 100% is absorbed by the atmosphere
- **D** 100% is reflected back to the Earth by the atmosphere

16 How is it possible that eaves can shade windows in the summer months, yet allow sun in during the winter months?

17 How can planting deciduous trees near windows contribute to good passive design?

18 Complete the following paragraph by by choosing the correct response from the choices given in brackets.

While fans don't actually [**increase/reduce**] the temperature or humidity, air flow increases [**convective/conductive**] cooling by moving air around the body, creating a similar effect to [**increasing/reducing**] the actual air temperature by around [**3°C/7°C**].

19 Explain why an electric-powered car may have higher emissions than a conventional car. Under what circumstances could an electric car be emission free?

UNIT 1 • Area of Study 1

REVIEW QUESTIONS

How can thermal effects be explained?

1 Three metals A, B and C are placed in thermal contact with one another. Heat flows from A to B and from C to B.

What can you say about the relative temperatures of metals A and C?

A A is at a higher temperature than C.
B A is at a lower temperature than C.
C A is at the same temperature as C.
D There is insufficient information to compare the temperatures of A and C.

2 Is it possible for two objects to be in thermal equilibrium if they are not in contact? Explain your answer.

3 Two objects A and B are at the same temperature. What does this imply about the kinetic energy of the molecules in each object?

A The kinetic energy of any molecule in object A will be the same as that of any molecule in object B.
B The average kinetic energy of the molecules in object A will be the same as in object B.
C The highest kinetic energy of any molecule in object A will be the same as in object B.
D The lowest kinetic energy of any molecule in object A will be the same as in object B.

4 A physics teacher plunges an inflated balloon into liquid nitrogen at 77 K. The balloon shrinks significantly and the students observe a small amount of liquid at the bottom of the balloon. When the balloon is taken out of the liquid nitrogen it slowly recovers its original volume and the liquid disappears. Explain the changes that the students observe in terms of the energy of the air particles in the balloon and the flow of energy involved.

5 A bucket is filled with equal amounts of hot and cold water. The hot water is originally at 80°C and the cold water at 10°C.

Hint: for all mixtures, heat is transferred from a hotter body to a cooler one until thermal equilibrium is reached. So here, $mc\Delta T_{\text{hot water}} = mc\Delta T_{\text{cold water}}$.

The temperature of the final mixture will be approximately:

A 10°C
B 45°C
C 70°C
D 90°C.

6 Which of the following examples supports the statement that it takes a larger amount of thermal energy to melt ice than to warm air?

A Moisture forms on the outside of a glass of ice water.
B Ice cubes in a freezer can be colder than 0°C.
C Glaciers and ice flows last throughout summer in some areas of New Zealand.
D Snow storms can occur at low altitudes in winter during extreme weather conditions.

The following information relates to questions 9 and 10.

Dry ice (solid CO_2) sublimes at –78.5°C to form gas at the same temperature. Sublimation is a direct change from a solid to a gas without a liquid phase in between.

7 Explain the change in internal energy that occurs when dry ice sublimes.

8 Explain why heat is absorbed without a temperature change.

9 A student attempts to identify a metal by measuring its specific heat capacity. A 100 g block of the metal is heated to 75°C and then transferred to a 70 g copper calorimeter containing 200 g of water at 20°C. The temperature of the final mixture is 25°C.

Using the table below and the hint in question 7, what metal is the student probably testing?

Material	Specific heat capacity (J kg^{-1} K^{-1})
human body	3500
methylated spirits	250
air	1000
aluminium	900
glass	840
iron	440
copper	390
brass	370
lead	130
mercury	140
ice	2100
liquid water	4200
steam	2000

10 A chemical engineer is doing a gas law calculation and understands that she needs to use temperature in kelvin. Her thermometer in the reactor vessel reads the temperature as 1550°C.

What temperature in kelvin would this be?

A 1277 K
B 1550 K
C 1823 K
D 2732 K

The following information relates to questions 11 to 13. When a copper rod at room temperature (25°C) is placed in a 1500°C furnace, heat flows into the rod.

11 Describe the flow in terms of the first law of thermodynamics.

12 Explain what happens to the kinetic energy of the metal atoms and relate this to the temperature of the metal rod.

13 The copper rod begins to glow bright red and is removed and placed on a steel plate in a cooling chamber filled with nitrogen gas. Explain the different means by which the copper rod can lose heat.

14 Boiling water results in steam formation at the same temperature. Which of the following has/have changed for the water molecules in the steam?

A average kinetic energy of the particles
B average potential energy of the particles
C total internal energy of the particles
D the total number of particles

The following information applies to questions 15 to 20. It is possible to make homemade ice cream by using a salt-water solution (brine) as a refrigerant for the cream and sugar mix. You can assume for the purposes of this exercise that the cream and sugar mix is 70% water. It is actually the water in the ice cream that freezes.

Salt is added to ice to make 5.00 kg of cold brine at −11°C. 500 g of cream and sugar mixture is cooled in a plastic bag which is plunged into the refrigerant.

Use the following data where appropriate:

Heat of fusion of water: 334 kJ kg^{-1}
Specific heat capacity of water: 4200 J $\text{kg}^{-1}\ \text{K}^{-1}$
Specific heat capacity of cream and sugar mix: 3.80×10^3 J $\text{kg}^{-1}\ {}^\circ\text{K}^{-1}$
Specific heat capacity of brine solution: 3.5×10^3 J $\text{kg}^{-1}\ {}^\circ\text{K}^{-1}$

15 Explain why the unit of the specific heat capacity is J $\text{kg}^{-1}\ {}^\circ\text{K}^{-1}$ but the heat of fusion has unit J kg^{-1}.

16 Calculate the amount of heat that needs to be extracted from the 500 g of cream and sugar at 25.0°C to reduce the mix to 0°C.

17 Explain why the specific heat capacity of the cream and sugar mix is less than that of water.

18 Once the mix starts to form ice crystals, the temperature doesn't drop any more. Explain why this is the case.

19 Calculate the heat that would need to be extracted for the mix at 0°C to completely freeze the 70% water in the mix.

20 Calculate the final temperature of the salt water when all the water in the ice cream is frozen, assuming no other heat loss.

The following information applies to questions 21 to 24. An entrepreneur is considering opening an ice cream parlour that makes ice cream using liquid nitrogen. The liquid nitrogen serves the dual purpose of freezing the water in the ice cream and aerating it at the same time. He intends to make ice cream that is 70% water. He assumes that only the water needs to freeze to set the ice cream.

He collects the following data:

Boiling temperature of liquid nitrogen at atmospheric pressure: 77.0 K
Specific heat capacity for nitrogen gas: 1.34 kJ $\text{kg}^{-1}\ {}^\circ\text{K}^{-1}$
Heat of vaporisation of liquid nitrogen: 199 kJ kg^{-1}
Specific heat capacity of cream and sugar mix: 3.80 kJ $\text{kg}^{-1}\ {}^\circ\text{K}^{-1}$
Heat of fusion of water: 334 kJ kg^{-1}

21 How much heat is required to vaporise 1.00 kg of liquid nitrogen at 77.0 K?

22 How much heat does 1.00 kg nitrogen gas absorb as it heats from 77.0 K to 0°C?

23 How much heat must be removed to cool 200 g of refrigerated sugar and cream mix at 8.00°C down to make ice cream at 0°C?

24 What mass of liquid nitrogen does the entrepreneur need per 200 g ice cream portion if he takes the sugar and cream mix from the fridge at 8.00°C and freezes it at 0°C?

25 On a scorching summer day a thirsty student finds a 500 mL bottle of lemonade in the boot of the car. She measures the temperature of the lemonade to be 32°C.

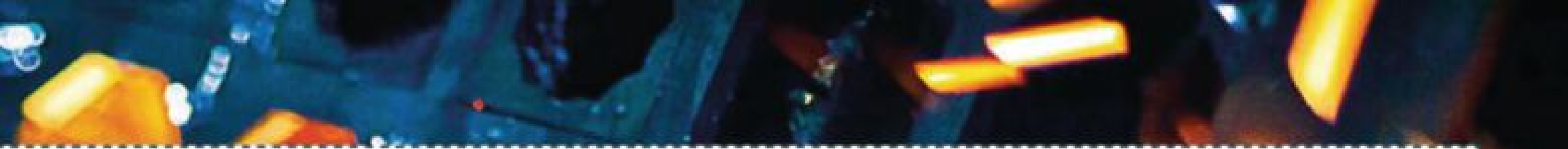
How much ice does she need to add to cool her lemonade down to a refreshing 4.0°C?

Assume the lemonade has the same properties as water.

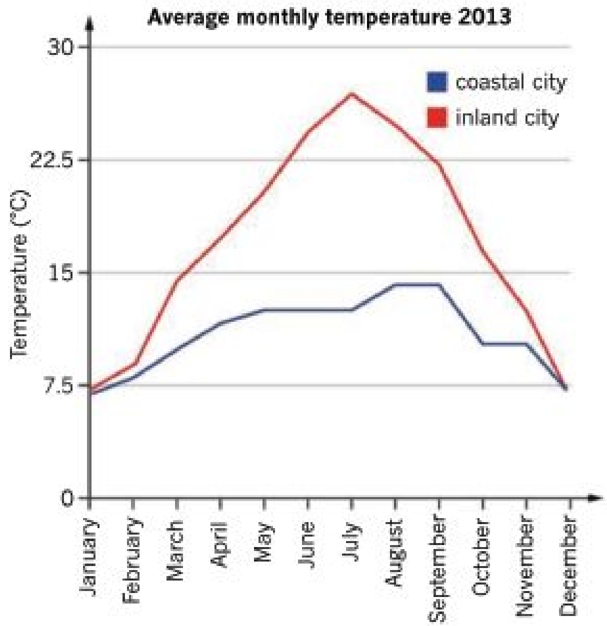
Hint: Watch out for unit conversions!

26 A 50.0 g minted coin is at 250.0°C when it is dropped into 500 g of water in a beaker at 20.0°C. If the final temperature of the system is 22.5 °C, calculate the Specific heat capacity of the metal.

Specific heat capacity of water: 4200 J $\text{kg}^{-1}\ \text{K}^{-1}$



27 The graph shows the comparative temperatures of two cities in the northern hemisphere at approximately the same latitude. The coastal city shows a much narrower range of temperature variation than the inland city. Use what you understand about water to explain why this is the case.



28 A patient burned by steam at 100°C sustains significantly more serious injuries than a patient who is scalded by boiling water. Propose an explanation for this.

29 According to the Stefan–Boltzmann law, the energy radiated by an ideal black body is proportional to the fourth power of the temperature in kelvin. If the average temperature of a bar radiator halves when it is switched from full to half power, the average transmission of thermal energy will:

A stay the same
B halve
C decrease 8 times
D decrease 16 times

30 A 65.0 kg patient suffering from a fever has a core temperature of 40.2°C and is placed in a cool bath to bring down her temperature. The specific heat capacity of a human is known to be 3.50 $\text{kJ kg}^{-1}\ °\text{K}^{-1}$. If the goal is to have the temperature of the patient and the bath reach 37.0°C, and the bath contains 40.0 L of water, what should the temperature of the bath water be?

1 L of water = 1 kg, and $c_{water} = 4200\ \text{J kg}^{-1}\ °\text{K}^{-1}$.

31 Order the following parts of the electromagnetic spectrum in terms of increasing:

a energy
b wavelength.

microwaves, infrared, X-rays, ultraviolet, radio waves, gamma rays, visible light

32 Describe the role of conduction, convection and radiation in moving heat around in the Earth's atmosphere.

Questions 33 to 37 refer to the energy balance of the Earth. The calculations carried out to evaluate global warming take into account the energy incident on the Earth from the Sun, and the energy that is re-radiated from the Earth. For an energy balance and stable global temperature, the energy input must equal the energy lost. Generally, energy incident on the Earth is counted as only occurring on the daylight side, while re-radiation occurs over the whole globe, and is calculated for an average Earth temperature.

33 The surface of Earth has an average temperature of 17°C. Assuming that the surface of Earth is a perfect radiator ($e = 1$), the power radiated out by each 1 m^2 of the Earth at this temperature would be (use $\sigma = 5.67 \times 10^{-8}$):

A 74×10^{-3} W
B 401 W
C 4.74×10^{5} W
D 40.1×10^{9} W

34 At the surface of the Earth, the solar power incident on one square metre is, on average, 1370 W m^{-2}. On average, the Earth reflects 31% of the incident power.

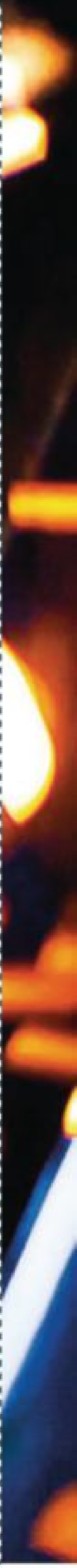
a Calculate the total power incident on the sunny side of the Earth, in terms of π and R_E (treating Earth as a disc of radius R_E).
b Calculate the power absorption of the side of the Earth exposed to the Sun, in terms of π and R_E.

35 Write an expression for the power radiated from the Earth at temperature T_E in terms of π and R_E over the whole surface of the Earth, assuming a constant temperature and using the Stefan–Boltzmann law.

36 The simple radiant input and reflection model does not take the greenhouse effect into account. How would the Earth's atmosphere alter the average temperature?

37 **a** A physics student who is also a keen bush camper notices that the night temperatures drop significantly lower on clear nights than on cloudy nights. Why is so?

b This same student is trying to calculate the effect of exposure on a hiker lost on a mountain for 5 h when the average outdoor temperature is –3°C. He estimates the hiker's surface temperature (outside his jacket) as 15 degrees. He calculates the radiant heat lost by the hiker in the 5 h. What should his answer be? Take the emissivity of the outer clothing to be 0.5 and the hiker's surface area to be 1.5 m^2.



38 A salesman at an outdoor shop is trying to sell a heater to a customer. He tells the customer that this heater, unlike the indoor ducted heating system, does not heat the air but rather heats objects. Is this just sales talk? Write a scientific explanation in terms of conduction, convection and radiation.

39 Simon puts a pot of peas in water to boil on the stove. The electric stove plate glows bright red. What heat transfer mechanisms cook the peas?

40 The surface temperature of the planets in the solar system drops as their distance from the Sun increases. Venus, with an atmosphere of 96.5% carbon dioxide, is a notable exception. Its temperature is much higher than would be expected given its distance from the Sun. Give a possible reason for this high temperature.

41 A planet called Proximus orbiting a star has a surface temperature of 600 K, while a more distant planet called Serus has a surface temperature of 300 K. What is the relationship between the power radiated per square metre from the planets?

42 Wien's law can be used to calculate the temperature of radiating objects. The Sun's temperature is 5778 K and the peak wavelength at which it radiates is about 500 nm. Calculate the temperature of the star Betelgeuse, which has maximum emission at 875 nm.

43 Military personnel use 'night-vision goggles' to detect enemy movement at night. How do these goggles work?

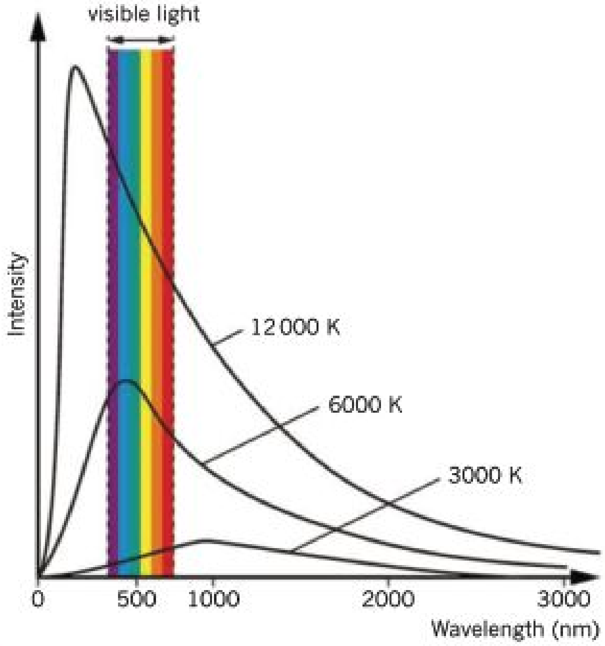
44 Wien's law can be written as $\lambda_{max}T = 2.898 \times 10^{-3}$ m K where λ is in m and the temperature is in K. Calculate:

a the peak radiation wavelength for the Earth at 290 K

b the temperature of a piece of steel glowing red hot in a furnace that radiates at 4.14 nm.

45 Consider your answer for the steel temperature above. 4.14 nm is not in the visible range of the spectrum. Explain why the steel is glowing red.

46 A metal filament is heated as a current passes through it. Initially it glows a dull red. As the current is increased, the red becomes brighter, the wire then glows yellow and eventually glows bright white. With reference to the following graph, explain the reasons for the colour changes of the metal filament at each stage.



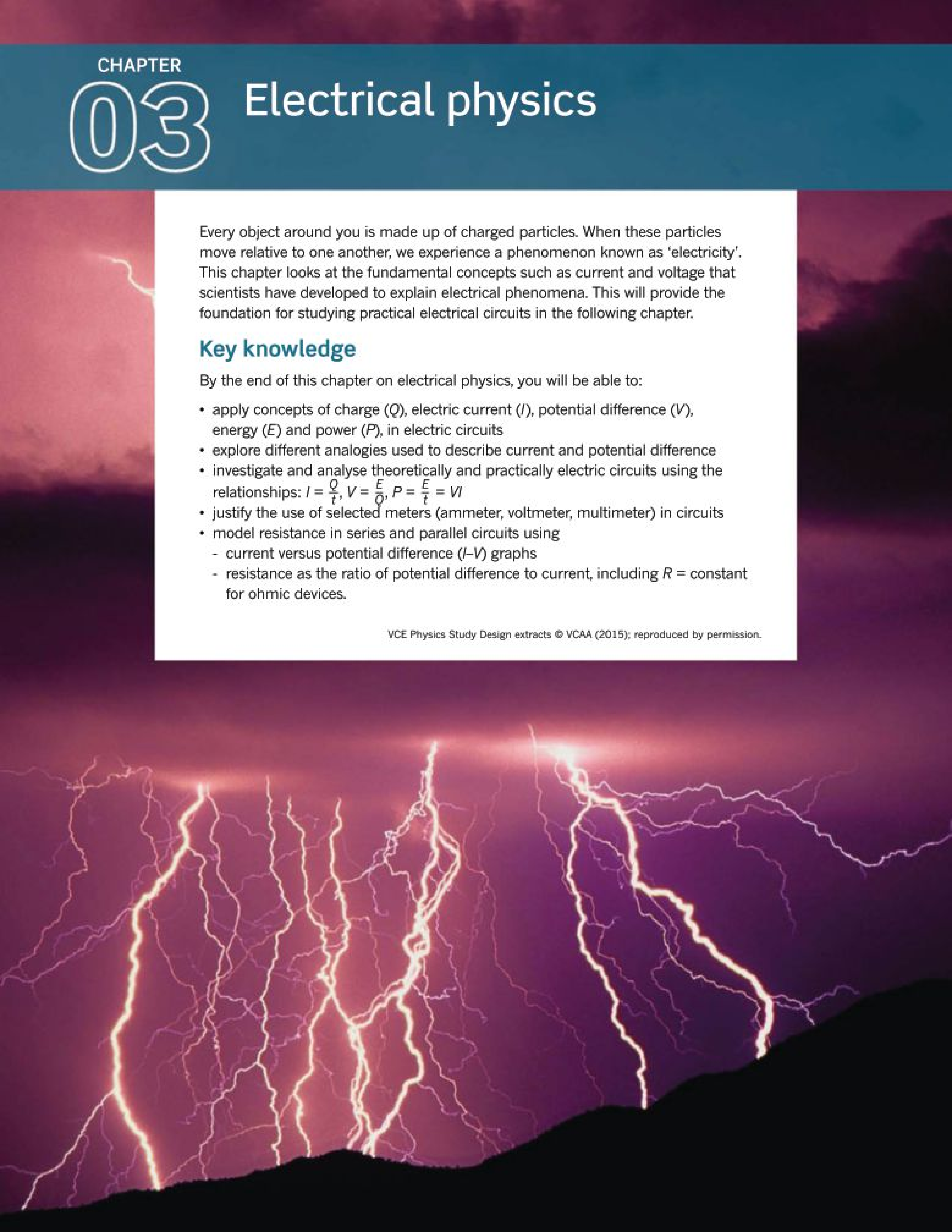
47 In the constellation Orion, Rigel is blue-white, while Betelgeuse is reddish. Based on this observation alone, what conclusion can be made?

A Rigel is further way than Betelguese.
B Rigel is closer than Betelguese.
C Rigel is cooler than Betelguese.
D Rigel is hotter than Betelguese.

48 Explain the difference between the greenhouse effect and the enhanced greenhouse effect.

49 Is there evidence for an increase in CO_2 levels in the atmosphere since the industrial age? Give examples.

50 Discuss some of the challenges facing scientists trying to interpret the data and decide whether global warming is directly attributable to human activity.



CHAPTER 03

Electrical physics

Every object around you is made up of charged particles. When these particles move relative to one another, we experience a phenomenon known as 'electricity'. This chapter looks at the fundamental concepts such as current and voltage that scientists have developed to explain electrical phenomena. This will provide the foundation for studying practical electrical circuits in the following chapter.

Key knowledge

By the end of this chapter on electrical physics, you will be able to:

- apply concepts of charge (Q), electric current (I), potential difference (V), energy (E) and power (P), in electric circuits
- explore different analogies used to describe current and potential difference
- investigate and analyse theoretically and practically electric circuits using the relationships: $I = \frac{Q}{t}$, $V = \frac{E}{Q}$, $P = \frac{E}{t} = VI$
- justify the use of selected meters (ammeter, voltmeter, multimeter) in circuits
- model resistance in series and parallel circuits using
  - current versus potential difference (I–V) graphs
  - resistance as the ratio of potential difference to current, including R = constant for ohmic devices.

3.1 Behaviour of charged particles

All matter in the universe is made of tiny particles. These particles have a property called **charge** that can be positive, negative or neutral. Usually, the numbers of positive and negative charges balance out so perfectly that we are completely unaware of them. However, when significant numbers of these charged particles become separated or move relative to each other, it results in **electricity**.

In order to understand electricity, it is important to first understand the way charged particles interact with each other.

EXISTENCE OF CHARGE CARRIERS

The tiny particles that make up all matter are called atoms. Every atom contains a nucleus at its centre. A nucleus is made up of positively charged particles called protons and neutral particles called neutrons. The nucleus, which is positively charged due to the protons, is surrounded by negatively charged electrons. A model of an atom is shown in Figure 3.1.1.

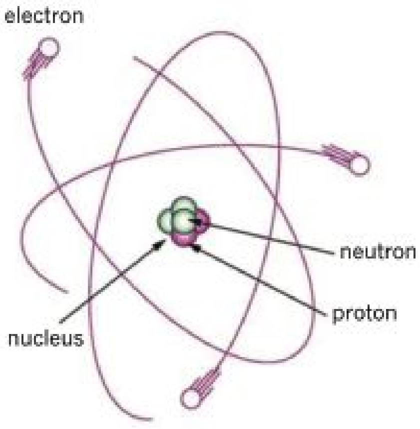


FIGURE 3.1.1 A simple model of an atom.

Simple models of the atom, often called planetary models, show the electrons as orbiting the nucleus much like the planets orbit the Sun, as you can see from the figure. This is because particles with like charges repel each other, but particles with opposite charges attract each other. In an atom the negatively charged electrons are attracted to the positively charged nucleus.

This is an important rule to remember when thinking about the interaction of charged particles.

Charge	Positive	Negative
Positive	repel	attract
Negative	attract	repel

PHYSICSFILE

Electron models

The way an electron moves around the nucleus of an atom is more complex than the simple planetary model would suggest. An individual electron is so small that its exact position at any point in time is impossible to measure. Recent models of the structure of the atom describe an electron in terms of the probability of finding it in a certain location. In diagrams of atoms, this is often represented by drawing the electrons around the nucleus as a fuzzy cloud, rather than points or solid spheres.

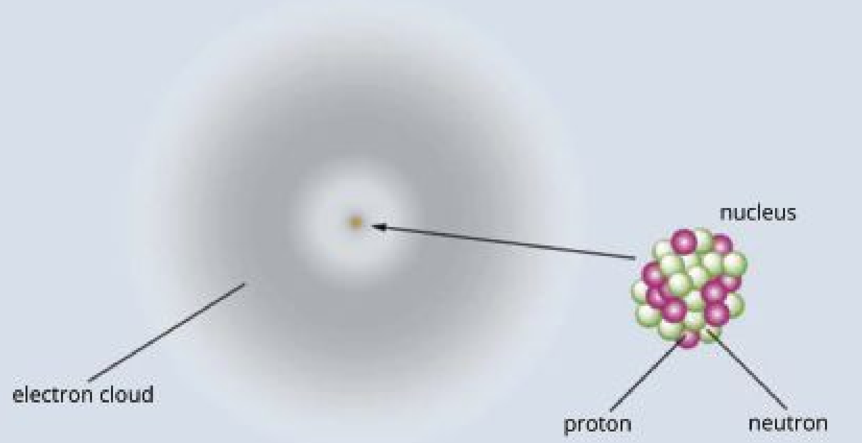


FIGURE 3.1.2 The nucleus of an atom occupies about 10^{-12} of the volume of the atom, yet it contains more than 99% of its mass. Atoms are mostly empty space.

In neutral atoms, the number of electrons is exactly the same as the number of protons. This means that their charges balance each other out, leaving the atom electrically neutral.

It is difficult to remove a proton from the nucleus of an atom. In comparison, electrons are loosely held to their respective atoms and it is relatively easy for them to be removed.

When electrons move from one object to another, each object is said to have gained a **net charge**. The object that loses the electrons will gain a net positive charge, since it will now have more positive protons than negative electrons. The object that gains electrons will gain a net negative charge. When an atom has gained or lost electrons, we say it has been **ionised** or has become an **ion**.

An excess of electrons causes an object to be negatively charged, and a deficit in electrons will mean the object is positively charged.

The understanding that the movement of electrons, rather than protons, creates electrical effects is a relatively new discovery. Unfortunately, this means that many of the rules and conventions used when talking about electricity refer to electric current as the movement of positive charge carriers.

PHYSICSFILE

Separating positive and negative charges

Electrons can be transferred (moved) from one object to another by simply rubbing two objects together. The objects need to be made from different materials. You can see this if you rub a balloon against your hair and then slowly move the balloon away. You will notice that your hair seems to stick to the balloon. This is because electrons are rubbed off your hair and transferred onto the balloon. This causes the balloon to gain a net negative charge and your hair to gain a net positive charge, which means the balloon and your hair are attracted to each other.

MEASURING CHARGE

In order to measure the actual amount of charge on a charged object, a 'natural' unit would be the magnitude (size) of the charge on one electron or proton. This fundamental charge is often referred to as the **elementary charge** and is given the symbol e. A proton therefore has a charge of $+e$ and an electron has a charge of $-e$.

The size of the elementary charge is very small. For most practical situations, it is more convenient to use a larger unit to measure charge. The SI (standard) unit of charge is known as the **coulomb** (symbol C). It is named after Charles-Augustin de Coulomb, who was the first scientist to measure the forces of attraction and repulsion between charges.

A coulomb is quite a large unit of charge: +1 coulomb (1 C) is equivalent to the combined charge of 6.2×10^{18} protons. Therefore, the charge on a single proton is $+1.6 \times 10^{-19}$ C. Similarly, −1 C is equivalent to the combined charge of 6.2×10^{18} electrons and the charge on a single electron is -1.6×10^{-19} C. The letter Q is used to represent the amount of charge.

The elementary charge, e, of a proton is equal to 1.6×10^{-19} C.
The elementary charge, $-e$, of an electron is equal to -1.6×10^{-19} C.

Worked example 3.1.1

THE AMOUNT OF CHARGE ON A GROUP OF ELECTRONS

Calculate the charge, in coulombs, carried by 6 billion electrons.	
Thinking	**Working**
Express 6 billion in scientific notation.	1 billion = 10^9 6 billion = 6×10^9
Calculate the charge, Q, in coulombs by multiplying the number of electrons by the charge on an electron (-1.6×10^{-19} C).	$Q = (6 \times 10^9) \times (-e)$ $= (6 \times 10^9) \times (-1.6 \times 10^{-19}\text{ C})$ $= -9.6 \times 10^{-10}$ C

Worked example: Try yourself 3.1.1

THE AMOUNT OF CHARGE ON A GROUP OF ELECTRONS

Calculate the charge, in coulombs, carried by 4 million electrons.

Worked example 3.1.2

THE NUMBER OF ELECTRONS IN A GIVEN AMOUNT OF CHARGE

The net charge on an object is −3.0 μC (1 μC = 1 microcoulomb = 10^{-6} C). Calculate the number of extra electrons on the object.	
Thinking	**Working**
Express −3.0 μC in scientific notation.	$Q = -3.0\ \mu\text{C}$ $= -3.0 \times 10^{-6}\ \text{C}$
Find the number of electrons by dividing the charge on the object by the charge on an electron (-1.6×10^{-19} C)	$n_e = \frac{Q}{-e} = \frac{-3.0 \times 10^{-6}\ \text{C}}{-1.6 \times 10^{-19}\ \text{C}}$ $= 1.9 \times 10^{13}$ electrons

Worked example: Try yourself 3.1.2

THE NUMBER OF ELECTRONS IN A GIVEN AMOUNT OF CHARGE

The net charge on an object is −4.8 μC (1 μC = 1 microcoulomb = 10^{-6} C). Calculate the number of extra electrons on the object.

ELECTRICAL CONDUCTORS AND INSULATORS

Electrons are much easier to move than protons. They also move more freely in some materials than in others.

In some materials, the electrons are only very slightly attracted to their respective nuclei. These materials are known as **metals**. Metals are good **conductors** of electricity. In conductors, loosely held electrons can effectively 'jump' from one atom to another and move freely throughout the material. This can be seen in Figure 3.1.3.

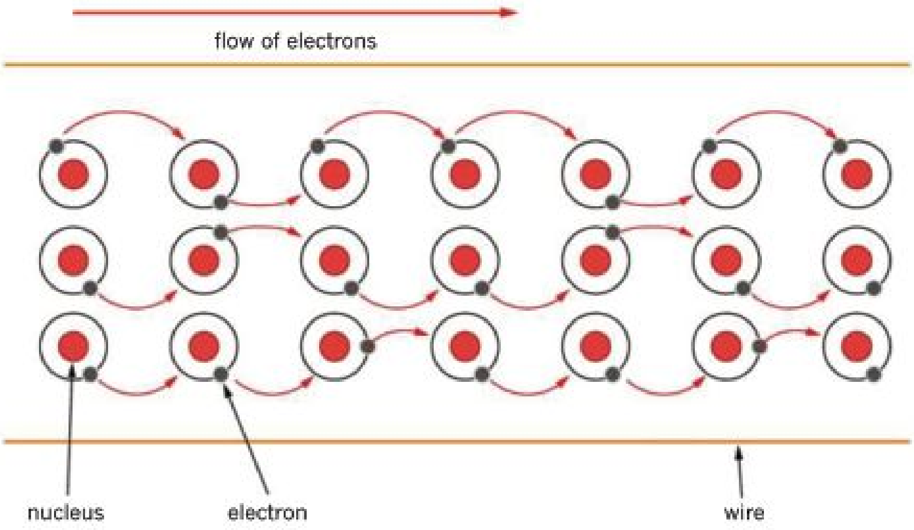


FIGURE 3.1.3 Electrons moving through a conductor. The electrons are free to move throughout the lattice of positive ions.

Copper is an example of a very good conductor. For this reason it is used in telecommunications and electrical and electronic products (see Figure 3.1.4).

FIGURE 3.1.4 These copper wires conduct electricity by allowing the movement of charged particles.

In comparison, the electrons in **non-metals** are very tightly bound to their respective nuclei and cannot readily move from one atom to another. Non-metals do not conduct electricity very well and are known as **insulators**. A list of common conductors and insulators is provided in Table 3.1.1.

Conductors	Insulators
Good	
all metals, especially silver, gold, copper and aluminium any ionic solution	plastics polystyrene dry air glass porcelain cloth (dry)
Moderate	
water earth semiconductors, e.g. silicon, germanium skin	wood paper damp air ice, snow

TABLE 3.1.1 Some common conductors and insulators.

PHYSICS IN ACTION

Lightning

Lightning is one of nature's greatest spectacles (see Figure 3.1.5). No wonder it was so long thought of as the voice of the gods. In the mid-eighteenth century, Benjamin Franklin showed that lightning is basically the same sort of electrical phenomenon as can be achieved by rubbing a glass rod with wool, or rubbing a balloon on your hair.

FIGURE 3.1.5 Lightning bolts over a city skyline.

A typical lightning bolt transfers 10 or more coulombs of negative charge (over 60 billion billion electrons) in approximately one thousandth of a second. A moderate thundercloud with a few flashes per minute generates several hundred megawatts of electrical power, the equivalent of a small power station.

It is thought that, during a thunderstorm, charge is transferred in collisions between the tiny ice crystals that form as a result of the cooling of upwards-flowing moist air and the larger, falling hailstones. As a result of small temperature differences between the crystals and hailstones, the crystals become positively charged and the hailstones negatively charged. The crystals carry their positive charge to the top of the cloud while the negative charge accumulates in the lower region. There is normally also a second smaller positively charged region at the bottom owing to positive charges attracted up from the ground towards the negative region, as seen in Figure 3.1.6.

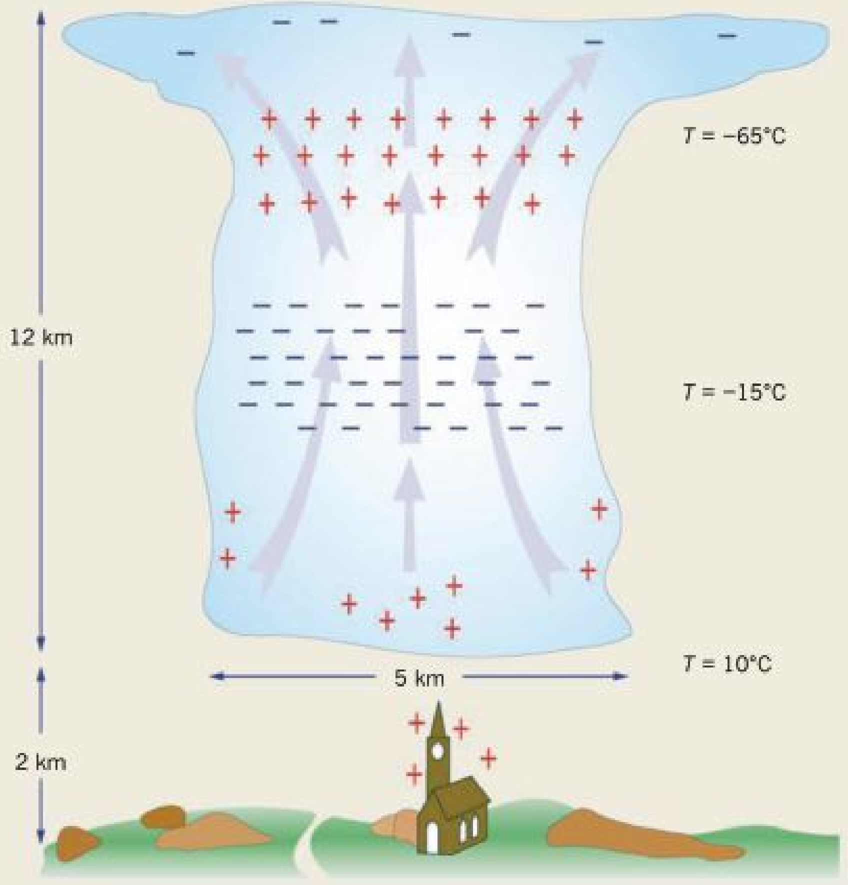


FIGURE 3.1.6 A thundercloud can be several kilometres wide and well over 10 km high. Strong updrafts drive the electrical processes that lead to the separation of charge. The strong negative charge of the lower region of the cloud will induce positive charges on tall objects on the ground. This may lead to a discharge, which can form a conductive path for lightning.

There will be strong electric fields between these regions of opposite charge. If they become sufficiently strong, electrons can be stripped from the air molecules (they become ionised). Because of the electric field, the free electrons and ions will gain kinetic energy and collide with more molecules, starting an 'avalanche of charges'. This is the lightning flash seen either within the cloud or between the Earth and the cloud. Most flashes are within the cloud; only a relatively small number actually strike the ground.

EXTENSION

Semiconductors

Some materials, such as silicon, are known as semimetals or metalloids. Their properties are somewhere between those of metals and non-metals. For example, the electrons in a silicon atom are not as tightly bound to the nucleus as those of a non-metal. However, they are not as easy to remove as the electrons in a metal. Hence, silicon and elements like it are known as semiconductors.

Silicon's ability to conduct electricity can be adjusted by adding small amounts of other elements such as boron, phosphorus, gallium or arsenic in a process known as doping. Adding another substance contributes free electrons, which can greatly increase the conductivity of silicon within electronic devices. This makes silicon very useful in the construction of computer chips like the one shown in Figure 3.1.7.

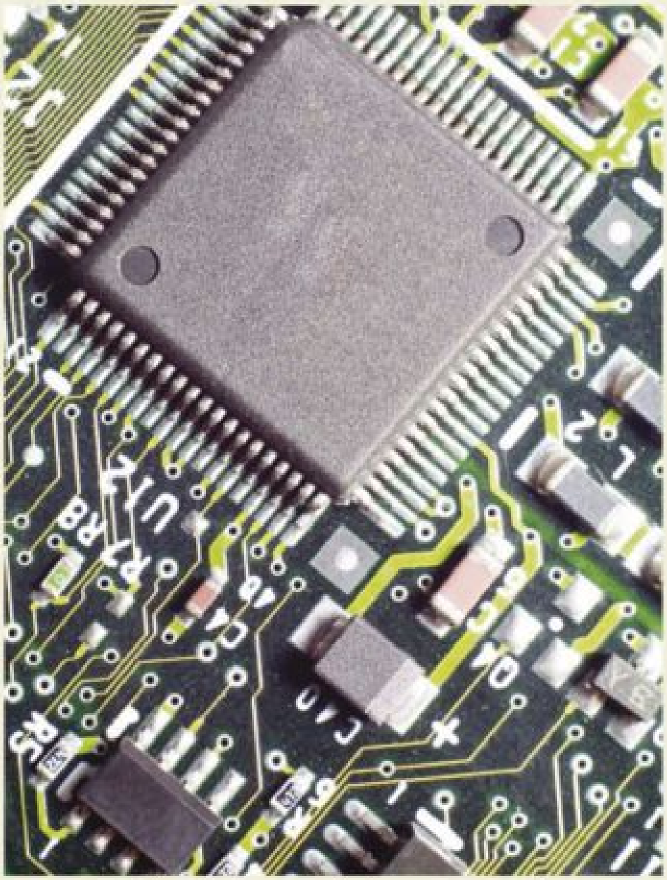


FIGURE 3.1.7 Silicon has been used in the construction of computer chips since the 1950s.

Much of the convenience of our modern lifestyle is based on the unique conductive properties of silicon and it has many electronic applications, some of which are shown in Figure 3.1.8.

FIGURE 3.1.8 All of these electronic devices use silicon in their construction.

3.1 Review

SUMMARY

- Like charges repel; unlike charges attract.
- When an object loses electrons, it develops a positive net charge; when it gains electrons, it develops a negative net charge.
- The letter *Q* is used to represent the amount of charge. The SI unit of charge is the coulomb (C).
- The elementary charge (*e*), the charge on a proton, is equal to 1.6×10^{-19} C. The elementary charge, $-e$, of an electron is -1.6×10^{-19} C.
- Electrons move easily through conductors, but not through insulators. This is because the electrons in materials that are good conductors are weakly attracted to the nucleus, and electrons in insulators are more strongly attracted to the nucleus.

KEY QUESTIONS

1 Plastic strip A, when rubbed, is found to attract plastic strip B. Strip C is found to repel strip B. What will happen when strip A and strip C are brought close together?

2 Calculate how many electrons make up a charge of −5.0 C.

3 Calculate the charge, in coulombs, of 4.2×10^{19} protons.

4 Explain why electric circuits often consist of wires that are made from copper and are coated in protective rubber.

3.2 Electric current and circuits

A flow of electric charge is called electric **current**. Current can be carried by moving electrons in a wire or by ions in solution. This section explores current as it flows through wires in electric circuits.

Electric circuits are involved in much of the technology used every day and are responsible for many familiar sights (see Figure 3.2.1). To construct electric circuits, you must know about the components of a circuit and be able to read circuit diagrams.

FIGURE 3.2.1 Electric circuits are responsible for lighting up whole cities.

ELECTRIC CIRCUITS

An **electric circuit** is a path made of conductive material, through which charges can flow in a closed loop. This flow of charges is called electric current. The most common conductors used in circuits are metals, such as copper wire. The charges that flow around the circuit within the wire are negatively charged electrons. The movement of electrons in the wire is called **electron flow**.

A simple example of an electric circuit is shown in Figure 3.2.2. The light bulb is in contact with the positive terminal (end) of the battery; a copper wire joins the **negative terminal** of the battery to one end of the filament in the light bulb. This arrangement forms a closed loop that allows electrons within the circuit to flow from the negative terminal towards the positive terminal of the battery. The battery is a source of energy. The light bulb converts (changes) this energy into other forms of energy, such as heat and light, when the circuit is connected.

If a switch is added to the circuit in Figure 3.2.2, the light bulb can be turned off and on. When the switch is closed, the circuit is complete. The current flows in a loop along a path made by the conductors and then returns to the battery.

When the switch is open, there is a break in the circuit and the current can no longer flow. This is what happens when you turn off the switch for a lamp or TV. A circuit where the conducting path is broken is often called an open circuit.

A switch on a power point or an appliance allows you to break the circuit. A break in the circuit occurs when two conductors in the switch are no longer in contact. This stops the flow of current and the appliance will not work.

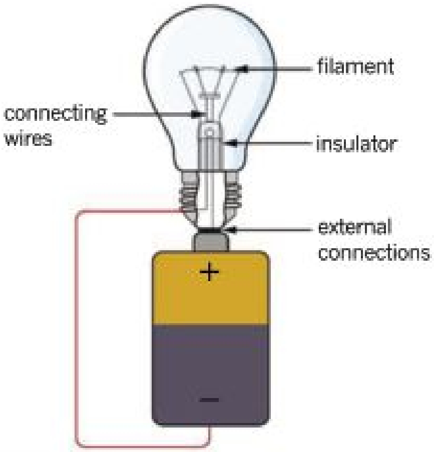


FIGURE 3.2.2 When there is a complete conduction path from the positive terminal of a battery to the negative terminal, a current flows.

Current will flow in a circuit only when the circuit forms a continuous (closed) loop from one terminal of a power supply to the other terminal.

REPRESENTING ELECTRIC CIRCUITS

Common symbols for electronic components

A number of different components can be added to a circuit. It is not necessary to be able to draw detailed pictures of these components; simple symbols are much clearer. The common symbols used to represent the electrical components in electric circuits are shown in Figure 3.2.3.

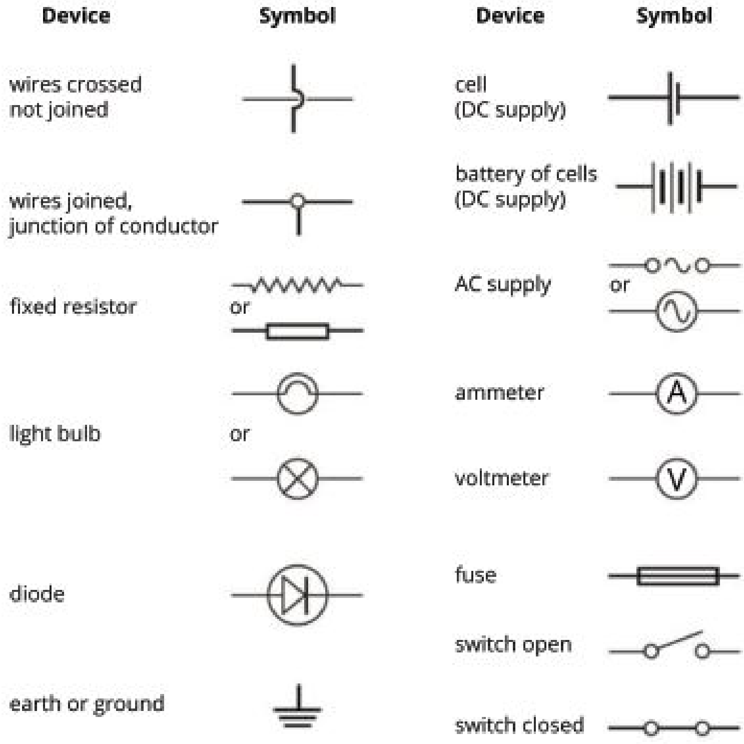


FIGURE 3.2.3 Some commonly used electrical devices and their symbols.

Circuit diagrams

When building anything, it is important that the builder has a clear set of instructions from the designer. This is as much the case for electric circuits as it is for a tall building or a motor vehicle.

Circuit diagrams are used to clearly show how the components of an electric circuit are connected. They simplify the physical layout of the circuit into a diagram that is recognisable by anyone who knows how to interpret it. You can use the list of common symbols for electrical components (Figure 3.2.3) to interpret any circuit diagrams.

The circuit diagram in Figure 3.2.4(b) shows how the components of the torch shown in Figure 3.2.4(a) are connected in a circuit. The circuit can be traced by following the straight lines representing the connecting wires.

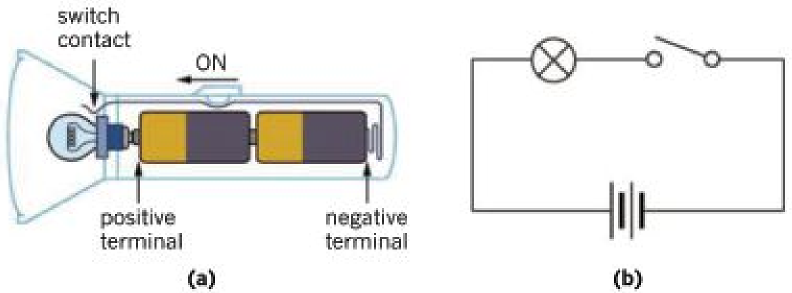


FIGURE 3.2.4 (a) A battery and light bulb connected by conductors in a torch constitute an electric circuit. (b) The torch's circuit can be represented by a simple circuit diagram.

Conventional current vs electron flow

When electric currents were first studied, it was (incorrectly) thought the charges that flowed in circuits were positive. Based on this, scientists traditionally talked about electric current as if current flowed from the positive terminal of the battery to the negative terminal. This convention is still used today, even though we know now that it is actually the negative charges (electrons) that flow around a circuit.

On a circuit diagram current is indicated by a small arrow and the symbol I. This is called **conventional current** or just current. The direction of conventional current is opposite to the direction of electron flow (Figure 3.2.5).

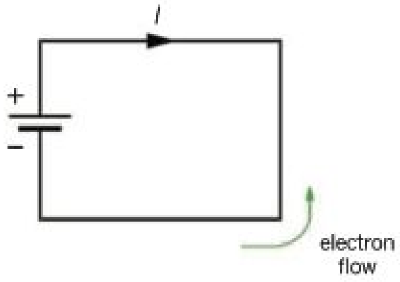


FIGURE 3.2.5 Conventional current (I) and electron flow are in opposite directions. The long terminal of the battery is positive.

> Conventional current (or current), I, flows from the positive terminal of a power supply to the negative terminal.
>
> Electron flow (or electron current) refers to the flow of electrons from the negative terminal to the positive terminal of a power supply.

QUANTIFYING CURRENT

One common misconception about current is that charges are used up or lost when a current flows around a circuit. However, the charge carriers (electrons) are conserved at all points in a circuit.

As current flows, electrons travel into the wire at the negative terminal of the battery. As electrons flow around a circuit, they remain within the metal conductor. They flow through the circuit and return to the battery at the positive terminal but are not lost in between.

In common electrical circuits, a current consists of electrons flowing within a copper wire (shown in Figure 3.2.6). This current, I, can be defined as the amount of charge that passes through a point in the conductor per second.

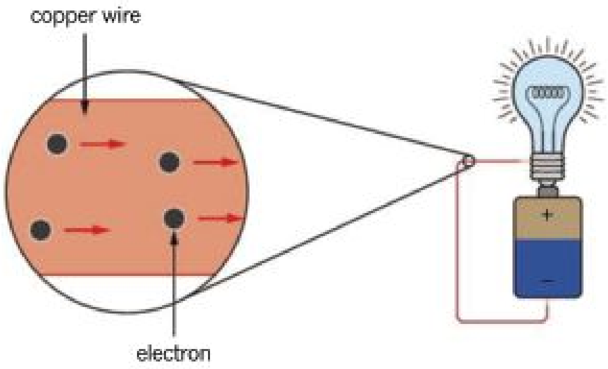


FIGURE 3.2.6 The number of electrons that pass through a point per second gives a measure of the current. Because the electrons do not leave the wire, current is conserved in all parts of the circuit.

> An equation to express this is:
>
> $I = \frac{Q}{t}$
>
> where I is the current in amperes
>
> Q is the amount of charge in coulombs
>
> t is the number of seconds that have passed.
>
> Current is measured in amperes, or amps (A). One ampere is equivalent to one coulomb per second.

PHYSICSFILE

Conventional current

Note that a flow of 1 C of *positive charge to the right* in one second is equivalent to a flow of 1 C of *negative charge to the left* in one second. That is, both situations represent a conventional current of 1 A to the right.

Current is a flow of electrons. It is equal to the number of electrons (n_e) that flow through a particular point in the circuit multiplied by the charge on one electron ($q_e = -1.6 \times 10^{-19}$ C) divided by the time that has elapsed in seconds (t). This makes the equation:

$$I = \frac{Q}{t} = \frac{n_e q_e}{t}$$

A typical current in a circuit powering a small DC motor would be about 50 mA. Even with this seemingly small current, approximately 3×10^{17} electrons flow past any point on the wire each second.

Worked example 3.2.1

USING $I = \frac{Q}{t}$

Calculate the number of electrons that flow past a particular point each second in a circuit that carries a current of 0.5 A.

Thinking	Working
Rearrange the equation $I = \frac{Q}{t}$ to make Q the subject.	$I = \frac{Q}{t}$ $\quad I \times t = \left(\frac{Q}{t}\right) \times t \quad I \times t = Q$ so $Q = I \times t$
Calculate the amount of charge that flows past the point in question by substituting the values given.	$Q = 0.5 \times 1$ $= 0.5$ C
Find the number of electrons by dividing the charge by the charge on an electron (1.6×10^{-19} C).	$n_e = \frac{Q}{q_e}$ $= \frac{0.5}{1.6 \times 10^{-19}}$ $= 3.12 \times 10^{18}$

Worked example: Try yourself 3.2.1

USING $I = \frac{Q}{t}$

Calculate the number of electrons that flow past a particular point each second in a circuit that carries a current of 0.75 A.

MEASURING CURRENT: THE AMMETER

Current is commonly measured by a device called an **ammeter** (see Figure 3.2.7).

Figure 3.2.7 shows the ammeter connected along the same path taken by the current flowing through the light bulb. This is referred to as connecting the ammeter 'in series'. Series circuits are covered in more detail in Chapter 4. The positive terminal of the ammeter is connected so that it is closest to the positive terminal of the power supply. The negative terminal of the ammeter is closest to the negative terminal of the power supply.

Measuring the current is possible because charge is conserved at all points in a circuit. This means that the current that flows into a light bulb is the same as the current that flows out of the light bulb. An ammeter can therefore be connected before or after the bulb in series to measure the current. Table 3.2.1 lists some typical values for electric current in common situations.

FIGURE 3.2.7 A digital ammeter (labelled with an A) measures current in a circuit.

Situation	Current
lightning	10 000 A
starter motor in car	200 A
fan heater	10 A
toaster	3 A
light bulb	400 mA
pocket calculator	5 mA
nerve fibres in body	1 μA

TABLE 3.2.1 Typical values for electric current.

ANALOGIES FOR ELECTRIC CURRENT

Since we cannot see the movement of electrons in a wire, it is sometimes helpful to use analogies or 'models' to visualise or explain the way an electric current behaves. It is important to remember that no analogy is perfect: it is only a representation and there will be situations where the electric current does not act as you would expect from the analogy.

Water model

A very common model is to think of electric charges as water being pumped around a pipe system, as shown in Figure 3.2.8. The battery pushes electrons through the wires just like a pump pushes water through the pipes. Since water cannot be compressed, the same amount of water flows in every part of a pipe, just as the electric current is the same in every part of a wire. Light bulbs in an electric circuit are like turbines: whereas the turbine converts the gravitational energy of the water into kinetic energy, a light bulb converts electrical energy into heat and light. The water that has flowed through the turbine flows back to the pump that provides the energy needed for it to keep flowing.

Comparing a DC circuit to the flow of water

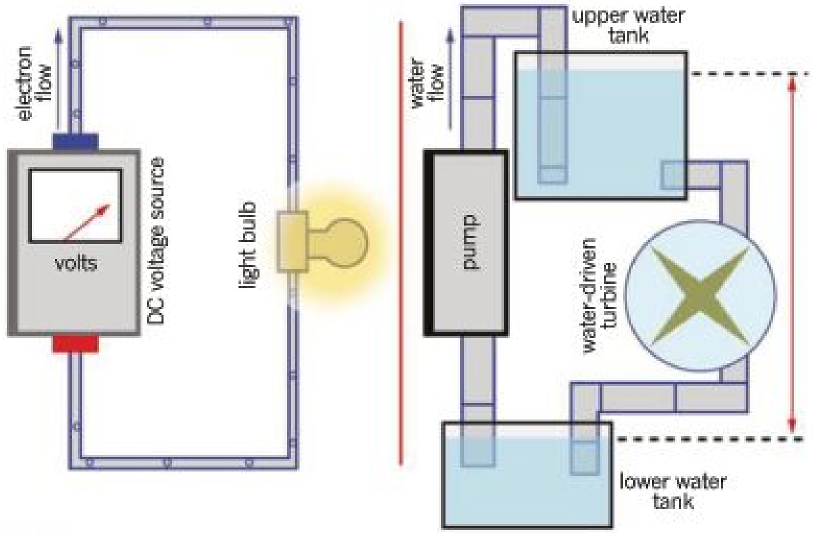


FIGURE 3.2.8 An electric current can be compared to water flowing through a pipe system.

This model explains the energy within a circuit quite well.

- The power supply transfers energy to the electrons and so the electrons gain potential energy.
- The energy of the electrons is converted into other forms when the electrons pass through the components in the circuit.

One of the limitations of the water model is that you usually cannot see water moving through a pipe and so you have to imagine what is happening in the pipe and then compare it to the motion of electrons in the wires.

PHYSICS FILE

Lightning power

Nature provides the most extreme examples of electric currents in the form of lightning. The movement of a huge number of charges heats a channel of air up to 30 000°C. The air then ionises and a current can flow. Some lightning bolts have currents greater than 200 000 A.

It is estimated that at any one time there are 2000 lightning storms around the globe. These create more than 100 lightning strikes every second. The energy from just one large thunderstorm would be enough to power all the homes in Australia for a few hours.

Bicycle chain model

Although electrons move relatively slowly through a conductor, electric effects are almost instantaneous. For example, the delay between flicking a light switch and the light coming on is too small to be noticed. One way of understanding this is to compare an electric current to a bicycle chain, as shown in Figure 3.2.9.

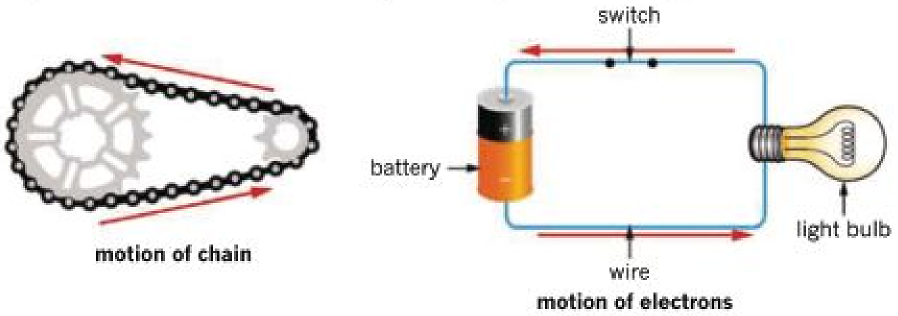


FIGURE 3.2.9 Electrons in a wire are like the links of a bicycle chain. Just like the links of a bicycle chain, electrons move together in a conductor.

Since a wire is full of electrons that all repel each other, moving one electron affects all the others around it. An electric current is like a bicycle chain: even if the cyclist pedals slowly, the links in the chain mean that energy is instantly transferred from the pedals to the wheel.

In this model, the pedals of the bicycle are like the battery of the electric circuit: the pedals provide the energy which causes the chain to move.

This model reinforces a number of important characteristics of an electric current.

- Electric effects are nearly instantaneous, just as there is no delay between turning the pedals and the back wheel of the bicycle turning.
- Charges in an electric current are not 'consumed' or 'used up', just as links in the chain are not used up.
- The amount of energy provided by an electric current is not entirely dependent on the current. This is like when a cyclist changes gears to give the same amount of energy to the bicycle while pedalling at different rates.

Although the bicycle chain model can be a helpful analogy, there are a number of important differences between a bicycle chain and an electric circuit.

- The number of charges flowing in an electric current is much larger than the number of links in a bicycle chain.
- Electrons in a wire do not touch one another like the links in a chain.

PHYSICSFILE

Electron flow

In any piece of conducting material, such as copper wire, electrons are present throughout the material. If there is no current flowing, this means there is no net flow of electrons, but the electrons are still present.

When you connect a piece of conducting material to the negative terminal of a battery, the negative terminal tries to 'push' the electrons away. However, the electrons will not flow if the circuit is open. This is because the electrons at the open part of the circuit have effectively reached a dead end, like cars stopped at a road block. This prevents all the other electrons in the material from flowing, like a long traffic jam caused by the road block. When you close the circuit, you create a clear pathway for the electrons to flow through. This means the electron closest to the negative terminal forces the next electron to move, and so on, all the way around the circuit. Therefore all electrons move almost simultaneously throughout the circuit so that electronic devices, such as light bulbs, seem to turn on immediately after you flick the switch.

3.2 Review

SUMMARY

- Current will flow in a circuit only when the circuit forms a continuous (closed) loop from one terminal of a power supply to the other terminal.
- When an electric current flows, electrons all around the circuit move towards the positive terminal, at the same time. This is called electron flow.
- Conventional current in a circuit flows from the positive terminal to the negative terminal.
- Current, I, is defined as the amount of charge, Q, that passes through a point in a conducting wire per second. It has the unit amperes or amps (A), which are equivalent to coulombs per second. The equation for this is:
  $I = \frac{Q}{t} = \frac{n_e q_e}{t}$
- Current is measured with an ammeter connected along the same path as the current flowing (that is, in series) within the circuit.

KEY QUESTIONS

1 What are the requirements for current to flow in a circuit?

2 List the electronic devices shown in the circuit diagram in Figure 3.2.10.

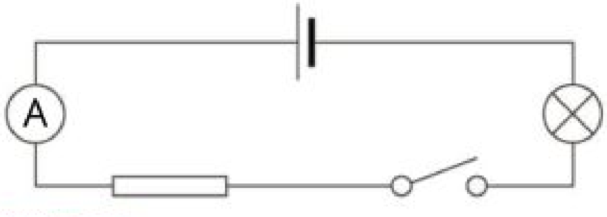


FIGURE 3.2.10

3 Why do scientists refer to conventional current as flowing from positive to negative?

A Protons flow from the positive terminal of a battery to the negative terminal.

B Electrons flow from the positive terminal of a battery to the negative terminal.

C Originally, scientists thought charge carriers were positive.

D Charges flow in both directions in a wire; conventional current refers to just one of the flows.

4 What current flows in a light bulb through which a charge of 30 C flows in:

a 10 seconds?

b 1 minute?

c 1 hour?

5 A car headlight may draw a current of 5 A. How much charge will have flowed through it in:

a 1 second?

b 1 minute?

c 1 hour?

6 **a** In a solution of salt water a total positive charge of +15 C moved past a point to the right in 5 s, and in the same time a total negative charge of −30 C moved to the left. What was the current through the solution during this time?

b Some time later it was found that in 5 s a total of +5 C had moved to the right while −15 C had moved to the right as well. What was the current this time?

7 Using the values given in Table 3.2.1, find the amount of charge that would flow through a:

a pocket calculator in 10 min

b car starter motor in 5 s

c light bulb in 1 h.

8 10^{20} electrons flow past a point in 4 seconds. Calculate:

a the amount of charge, in coulombs, that moves past a point in this time

b the current, in amps.

9 3.2 C flow past a point in 10 seconds. Calculate:

a the number of electrons that move past a point in this time

b the current, in amps.

10 Which of the circuits shown in Figure 3.2.11, below, would enable you to measure the current passing through both light bulbs when the switch is closed?

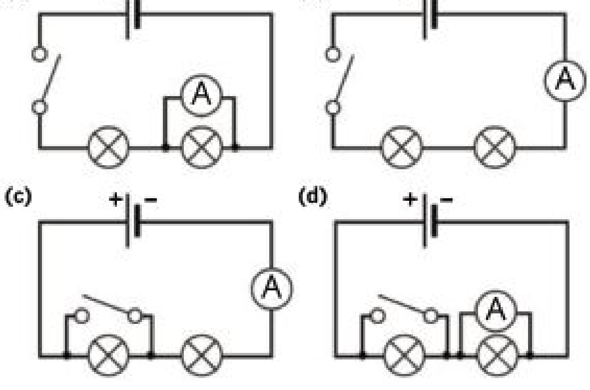


FIGURE 3.2.11

3.3 Energy in electric circuits

Electrons won't move around a circuit unless they are given energy. This energy is given by the battery (Figure 3.3.1). This is because inside every battery, a chemical reaction is taking place. The chemical reactions provide potential energy to the electrons inside it. When a circuit connects two ends of the battery, the potential energy of the electrons is converted into kinetic energy and the electrons move through the wire.

FIGURE 3.3.1 Chemical energy is stored within batteries.

Chemical reactions in the battery drive electrons towards its negative terminal. The electrons at the negative terminal repel each other. This repulsion moves them into the wire. At the positive terminal, electrons in the wire are attracted to the positive charges created by the deficiency of electrons. This attraction causes them to move into the battery. The net effect of electrons flowing into the wire at one end and out of it at the other end is that an electric current flows through the wire.

ENERGY IN CIRCUITS

Chemical energy stored inside a battery is **transformed** (changed) into **electrical potential energy**. This potential energy is stored as a separation of charge between the two terminals of the battery. This can be visualised as a 'concentration' of charge at either end of the battery. One terminal (the negative terminal) has a concentration of negative charges; the other terminal (the positive terminal) has a concentration of positive charges. Once the battery is connected within a device, chemical reactions will, for some time, maintain this difference in charge between the two terminals.

The difference in charge between the two terminals of a battery can be quantified (given a numerical value) as a difference in the electrical potential energy per unit charge. This is commonly called **potential difference** (V) and is measured in volts (V).

It is this potential difference at the terminals of the battery that provides the energy to a circuit. The energy is then **transferred** (passed) to different components in the circuit. At each component the energy is transformed into a different type of energy. For example, the energy could be transformed into heat and light if the component is a light bulb. If the component is a fan, the energy is transformed into motion (kinetic energy) and some heat and sound.

PHYSICSFILE

Cells and batteries

A single cell generates electricity by converting chemical energy to electrical potential energy. If a series of cells are added together, it is called a battery. Often a series of cells are packaged in a way that makes it look like a single device, but inside is a battery of cells connected together (see Figure 3.3.2). The terms 'battery' and 'cell' can be used interchangeably as the term 'battery' is frequently used in common language to describe a cell.

FIGURE 3.3.2 A mobile-phone battery. The term battery actually refers to a group of electric cells connected together.

PHYSICSFILE:

If we use a conductor to link two bodies between which there is a potential difference, charges will flow through the conductor until the potential difference is equal to zero (see Figure 3.3.3). For the same reasons, when a conductor is charged, charges will move through it until the potential difference between any two points in the conductor is equal to zero.

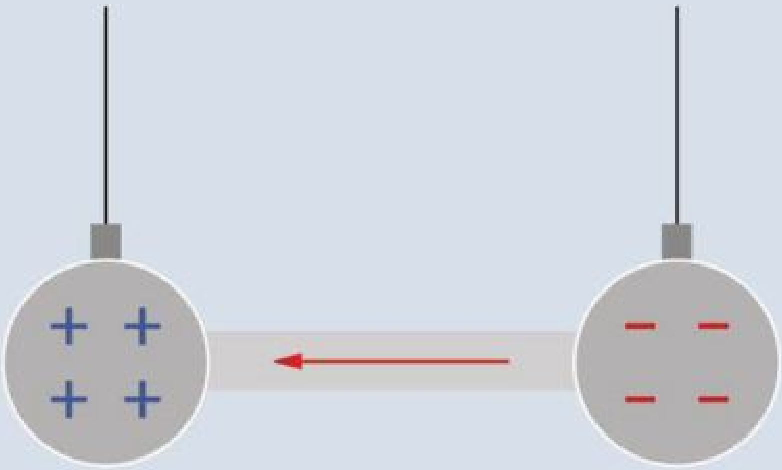


FIGURE 3.3.3 A potential difference exists between these two objects due to the difference in charge concentration. Electrons flow from the negative object to the positive object, as shown by the arrow, until the potential difference is zero.

Energy transfers and transformations in a torch

A torch is a simple example of how energy is transformed and transferred within a circuit. In the torch shown in Figure 3.3.4, chemical energy in the battery is transformed to electrical potential energy. There are two batteries connected so a bigger potential difference is available. Energy can be transferred to the light bulb once the end terminals of the batteries have been connected to the torch's circuit: that is, when the torch is switched on.

The electrical potential difference between the battery's terminals causes electrons within the circuit to move. The electrons flow through the wires of the torch. These electrons collide with the atoms within the small wire (filament) in the torch's light bulb and transfer kinetic energy to them. This transfer of kinetic energy means that the particles inside the filament move faster and faster and the filament gets very hot. When hot, the filament emits visible light.

The energy changes can be summarised as:

chemical energy $\xrightarrow{\text{transformed}}$ electrical potential energy
electrical potential energy $\xrightarrow{\text{transformed}}$ kinetic energy (electrons)
kinetic energy (electrons) $\xrightarrow{\text{transferred}}$ kinetic energy (filament atoms)
kinetic energy (filament atoms) $\xrightarrow{\text{transformed}}$ thermal energy + light

Eventually, when most of the chemicals within the battery have reacted, the battery is no longer able to provide enough potential difference to power the torch. This is because the chemical reaction has slowed and electrons are not being driven to the negative terminal. The torch stops working and the batteries are said to have gone flat.

Similar energy transfers and transformations take place every time electrical energy is used.

PHYSICSFILE

Volts and voltage

Somewhat confusingly, scientists use the symbol 'V' for *both* the quantity potential difference *and* its unit of measurement, the volt. For the quantity potential difference, we use italics: *V* For the unit volts, the symbol is not in italics: V. The context usually makes it clear which meaning is intended.

For this reason, potential difference is often referred to as 'voltage'.

FIGURE 3.3.4 An X-ray image of the internal structure of a torch. The bulb and two batteries are clearly visible.

EXPLAINING POTENTIAL DIFFERENCE

When charges are separated in a battery, each charge gains electrical potential energy. In a similar way, if a mass is lifted above the ground it gains gravitational potential energy. The *change* in electrical potential energy of each charge is known as the potential difference (V).

As you can see in Figure 3.3.5, when you lift an object to some height above the ground, you have done some work and have placed it in the field where it has more gravitational potential energy (E_g) available to it.

The changing potential energy of a moving mass in a gravitational field

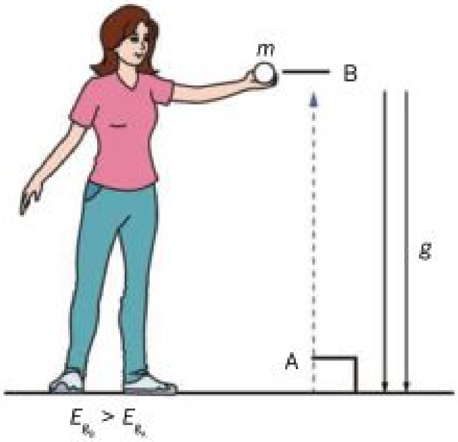


FIGURE 3.3.5 Lifting a mass above ground level increases its gravitational potential energy just as moving an electron to the negative terminal of an electric cell gives it electrical potential energy.

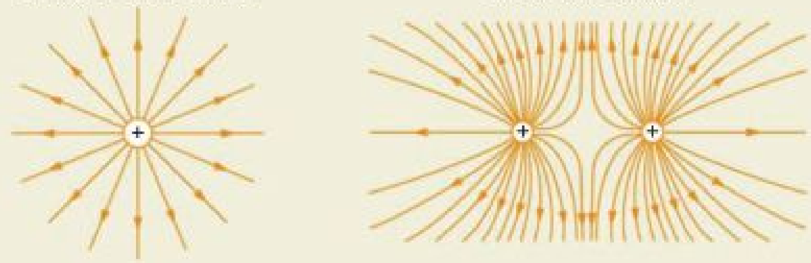
EXTENSION

Fields

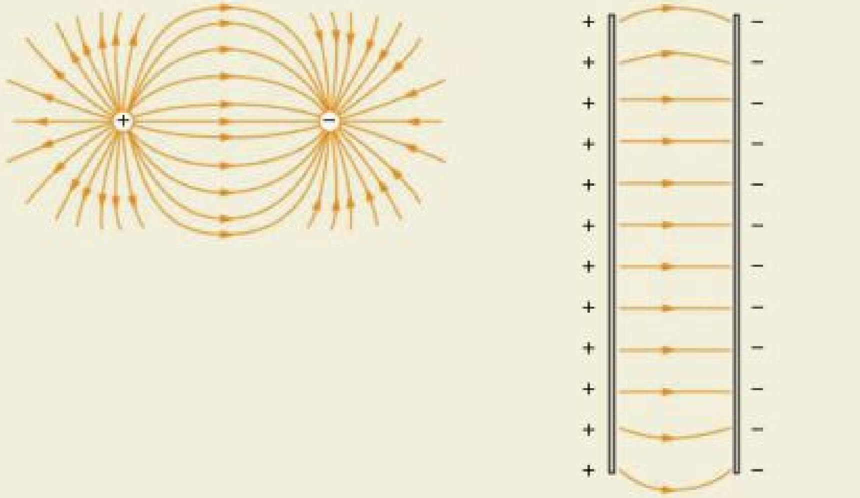
Just as the region around a massive object like the Earth contains a 'gravitational field' so too can the region inside a battery or electric cell be described as an 'electric field'. The idea of a field is useful for explaining how forces can act at a distance.

The electric field, like the gravitational field, is the amount and direction of the force on one unit—one unit of (positive) charge in the case of the electric field, one unit of mass in the case of the gravitational field.

All charged bodies produce an electric field around themselves. The field is represented by the so-called electric field lines, or lines of force. These lines indicate the direction of the force that would be applied by the field on a positive charge placed in the field. The shape of the field around some charged objects is shown in Figure 3.3.6. The concept of fields will be developed further in Unit 3.


a single positive charge

two positive charges


a positive and a negative charge

parallel oppositely charged plates

FIGURE 3.3.6 Electric fields can be represented by lines showing the direction of the field. The closeness of the lines suggests the strength of the field.

PHYSICSFILE

The term potential difference can often cause confusion. Stated simply, it means that there is a difference in electrical potential energy per charge between two points. The potential difference of a battery refers to the difference in electrical potential energy per charge at either terminal of the battery. You may see potential difference used to describe what happens on either side of a component, such as a light globe. If you measure the potential energy of charges on one side of a globe and compare it to the potential energy of the charges on the other side, you would find that there is a difference. The difference equates to the energy transferred from the charges to the globe to light it up.

PHYSICSFILE

Birds on a wire

Birds can sit on power lines and not get electrocuted even though the wires are not insulated.

For a current to flow through a bird on a wire, there would have to be a potential difference between its two feet. Since the bird has both feet touching the same wire, which might be at a very high potential (voltage), there is no potential *difference* between the bird's feet. If the bird could stand on the wire and touch any other object such as the ground or another wire, then it would get a big electric shock. This is because there would be a potential difference between the wire and the other object and current would flow.

FIGURE 3.3.7 There is no potential difference between each bird's feet.

Quantifying potential difference: voltage

As with other forms of energy, it is useful to be able to quantify the amount of potential difference in a given situation. Potential difference is formally defined as the amount of electrical potential energy given to each coulomb of charge. As an equation, it is:

$$V = \frac{E}{Q}$$

where V is potential difference (V)

E is electric potential energy (J)

Q is charge (C).

Since energy is measured in joules and charge in coulombs, the potential difference is measured in joules per coulomb (J C^{-1}). This quantity has been assigned a unit, the **volt** (V). So a potential difference of $1\ \text{J C}^{-1}$ is equal to 1 V.

PHYSICSFILE

Energy in a battery

When a battery is labelled 9 V, this means that the battery provides 9 J of energy to each coulomb of charge.

Worked example 3.3.1

DEFINITION OF POTENTIAL DIFFERENCE

Calculate the amount of electrical potential energy carried by 5 C of charge at a potential difference of 10 V.

Thinking	Working
Recall the definition of potential difference.	$V = \frac{E}{Q}$
Rearrange this to make energy the subject.	$E = VQ$
Substitute in the appropriate values and solve.	$E = 10 \times 5$ $= 50$ J

Worked example: Try yourself 3.3.1

DEFINITION OF POTENTIAL DIFFERENCE

A car battery can provide 3600 C charge at 12 V. How much electrical potential energy is stored in the battery?

Measuring voltage: the voltmeter

Voltage is usually measured by a device called a **voltmeter**.

Voltmeters are wired into circuits differently to ammeters. Unlike an ammeter, which measures the current passing through a wire, a voltmeter measures the change in voltage (potential difference) as current passes through a particular component. This means that one wire of the voltmeter is connected to the circuit before the component and the other wire is connected to the circuit after the component. This is called connecting the voltmeter 'in parallel', making a **parallel circuit**.

In Figure 3.3.8, the voltmeter is connected to the circuit either side of the light globe, that is, in parallel. This is so that it can measure the voltage drop (potential difference) across the light globe. Like the ammeter, it is important to connect the voltmeter with the positive terminal closest to the positive terminal of the power supply. The voltmeter's negative terminal is connected closest to the negative terminal of the power supply.

FIGURE 3.3.8 A voltmeter measures the voltage change (in this case, 6.23 V) across a light globe.

QUANTIFYING ELECTRICAL ENERGY

Work done by a circuit

In electrical circuits, electrical potential energy is converted into other forms of energy. When energy is changed from one form to another, work is done. (Work is covered in more detail in Chapter 12.) The amount of energy provided by a particular circuit can be calculated using the definitions for potential difference $V = \frac{E}{Q}$ and current $I = \frac{Q}{t}$.

Rearranging the definition of voltage gives:

$E = VQ$

Using the definition of current:

$Q = It$

Therefore:

$E = VIt$

where E is the energy provided by the current, which is the same as the work done (J)

V is the potential difference (V)

I is the current (A)

t is the time (s).

This gives us a practical way to calculate the energy used in a circuit from measurements we can make.

Worked example 3.3.2

USING $E = VIt$

A potential difference of 12 V is used to generate a current of 750 mA to heat water for 5 minutes. Calculate the energy transferred to the water in that time.	
Thinking	**Working**
Convert the quantities to SI units.	$\frac{750\text{ mA}}{1000} = 0.750\text{ A}$ $5\text{ min} \times 60\text{ s} = 300\text{ s}$
Substitute values into the equation and calculate the amount of energy in joules.	$E = VIt$ $= 12 \times 0.750 \times 300$ $= 2700\text{ J}$

Worked example: Try yourself 3.3.2

USING $E = VIt$

A potential difference of 12 V is used to generate a current of 1750 mA to heat water for 7.5 minutes. Calculate the energy transferred to the water in that time.

Rate of doing work: power

If you wanted to buy a new kettle, you might wonder how you could determine how quickly different kettles boil water.

Printed on all appliances is a rating for the **power** of that device. Power is a measure of how fast energy is converted by the appliance. In other words, power is the rate at which energy is transformed by the components within the device. This can also be described as the rate at which work is done. As an equation:

$$P = \frac{\text{energy transformed}}{\text{time}} = \frac{E}{t}$$

where P is the power in joules per second (J s^{-1}). One joule per second is 1 watt (W).

PHYSICSFILE

Multimeters

The internal circuitry of a voltmeter is significantly different to an ammeter. Electricians and scientists who work with electrical circuits often find it inconvenient to keep collections of voltmeters and ammeters so they have found a way to bundle the circuitry for both meters up into a single device known as a *multimeter*. A multimeter is shown in Figure 3.3.9.

This is much more convenient because the multimeter can be quickly switched from being an ammeter to being a voltmeter as needed. However, when a multimeter is switched from one mode to another, it is important to make a corresponding change to the way it is connected to the circuit being measured. An ammeter is connected in series and a voltmeter in parallel. In fact, if a multimeter is working as an ammeter and it is connected in parallel like a voltmeter, it may draw so much current that its internal circuitry will be burnt out and the multimeter will be destroyed.

FIGURE 3.3.9 A digital multimeter can be used as either an ammeter or a voltmeter.

The more powerful an appliance is, the faster it can do a given amount of work. In other words, an appliance that draws more power can do the same amount of work in a shorter amount of time. If you want something done quickly, then you need an appliance that has a higher power rating.

Rearranging the previous relationship:

$E = VIt$ to $\frac{E}{t} = VI$

Combining this with the power expression gives you:

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

This expression enables us to calculate the energy transformations in a circuit by measuring voltage and current across circuit components. The power dissipated by those components can be calculated in watts (W).

Worked example 3.3.3

USING $P = VI$

An appliance running on 230 V draws a current of 4 A. Calculate the power used by this appliance.	
Thinking	**Working**
Identify the relationship needed to solve the problem.	$P = VI$
Identify the required values from the question, substitute and calculate.	$P = VI$ $= 230 \times 4$ $= 920$ W

Worked example: Try yourself 3.3.3

USING $P = VI$

An appliance running on 120 V draws a current of 6 A. Calculate the power used by this appliance.

ANALOGIES FOR POTENTIAL DIFFERENCE

The analogies used for electric current in Section 3.2 'Electric current and circuits' can also be used to understand the concept of potential difference.

In the water model (p. 99), potential difference is similar to the water pressure in the pipe. If the water is pumped into a raised water tank like in Figure 3.2.8 on page 99, potential difference can also be compared to the gravitational potential energy given to each drop of water.

In the bicycle chain analogy (p. 100), potential difference is related to how hard the bicycle is being pedalled. If the cyclist is pedalling hard, this would correspond to a high voltage in which each link in the chain is carrying a larger amount of energy than if the cyclist was pedalling slowly.

In both analogies the overall rate of energy output—that is, the power—is related to both the current and the potential difference. In the water analogy, the pressure in the pipe could be very high but the rate of energy transfer will depend on how quickly the water is flowing. Similarly, a cyclist can work at the same rate by pedalling hard with the chain moving slowly or pedalling more easily but with the chain moving more quickly.

3.3 Review

SUMMARY

- Electric potential difference measures the difference in electric potential energy available per unit charge
- Potential difference can be defined as the work done to move a charge against an electric field between two points, using the equation:
  $E = VQ$ or $V = \frac{E}{Q}$
- In a circuit, the energy required for charge separation is provided by a cell or battery. The chemical energy within the cell is transformed into electric potential energy.
- Power is the rate at which energy is transformed in a circuit component. It is defined and quantified by the relationships:
  $P = \frac{E}{t} = VI$

KEY QUESTIONS

1 Under what conditions will charge flow between two bodies linked with a rod? Choose the correct response from the following options.
  A The potential difference between the bodies is not zero and the rod is made of a conducting material.
  B The potential difference between the bodies is not zero and the rod is made of an insulating material.
  C The potential difference between the bodies is equal to zero and the rod is made of a conducting material.
  D The potential difference between the bodies is equal to zero and the rod is made of an insulating material.

2 A freezer has a power rating of 460 W and it is designed to be connected to 230 V. Calculate:
  a the work performed by the freezer in 5 minutes
  b the current flowing through the freezer.

3 **a** What is the potential of a battery that gives a charge of 10 C:
    i 40 J of energy in 1 second?
    ii 40 J of energy in 10 seconds?
    iii 20 J of energy in 10 seconds?
  b What current flowed in each case?

4 A charge of 5 C flows from a battery through an electric water heater and delivers 100 J of heat to the water. What was the potential difference of the battery?

5 How much charge must have flowed through a 12 V car battery if 2 kJ of energy was delivered to the starter motor?

6 A light bulb that is connected to 240 V uses 3.6 kJ of electric potential energy in one minute.
  a Into what type(s) of energy has the electrical energy been transformed?
  b Calculate the power of the lamp.
  c Calculate the current flowing through the lamp.

7 In comparing the electrical energy obtained from a battery to the energy of water stored in a hydroelectric dam in the mountains, to what could the potential difference of the battery be likened?

8 Andy wishes to measure the current and potential difference for a light bulb. He has set up a circuit as shown in Figure 3.3.10.

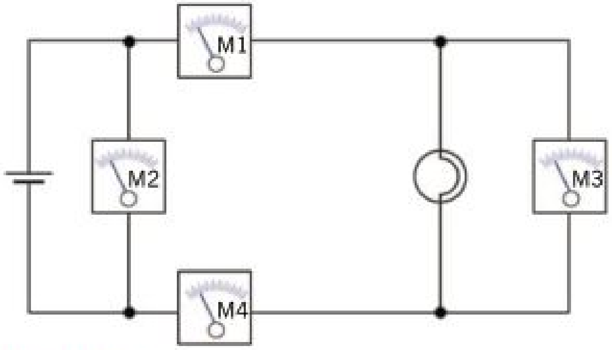


FIGURE 3.3.10

In which positions (M1, M2, M3 or M4) can he place:
  a a voltmeter?
  b an ammeter?

3.4 Resistance

Resistance is an important concept because it links the ideas of potential difference and current. **Resistance** is a measure of how hard it is for current to flow through a particular material. As conductors allow current to pass through easily, they are said to have low resistance. Insulators have a high resistance because they 'resist' or limit the flow of charges through them.

For a particular object or material, the amount of resistance can be quantified (given a numerical value). This means that the performance of electrical circuits can be studied and predicted with a high degree of confidence.

- Resistance is a measure of how hard it is for current to flow through a particular material.
- Resistance is measured in ohms (Ω).

RESISTANCE TO THE FLOW OF CHARGE

Energy is required to create and maintain an electric current. For electrons to move from one place to another, they need to first be separated from their atoms and then given energy to move. In some materials (i.e. conductors), the amount of energy required for this is negligible (almost zero). In insulators, a much larger amount of energy is required.

Once the electrons are moving through the material, energy is also required to keep them moving at a constant speed. Consider an electron travelling through a piece of copper wire. It is common to imagine the wire as an empty pipe or hose through which electrons flow. However, a piece of copper wire is not empty—it is full of copper ions. These ions are packed tightly together in a lattice arrangement. As an electron moves through the wire, it will 'bump' into the ions. The electron needs constant 'energy boosts' to keep it moving in the right direction. This is why an electrical device will stop working as soon as the energy source (e.g. battery) is disconnected.

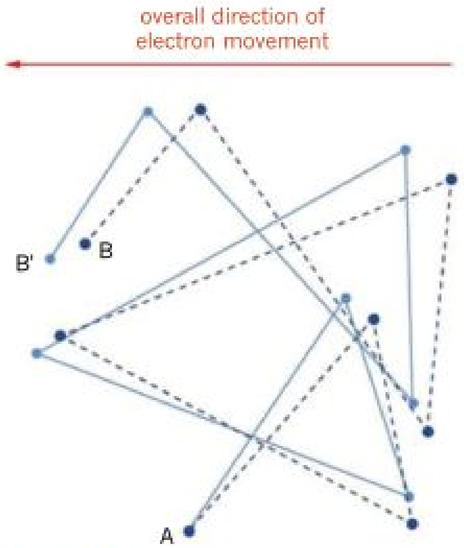


FIGURE 3.4.1 Path AB shows the random motion of an electron due to thermal effects. Path AB' shows the path of the same electron when an electric current is flowing in the direction indicated.

PHYSICSFILE

Electron movement

Even when current is not flowing, free electrons tend to move around a piece of metal due to thermal effects. The free electrons are rushing around at random with great speed. The net speed of an electron through a wire, however, is quite slow. Figure 3.4.1 compares the random motion of an electron when current is not flowing (AB) to the motion of the electron when current is flowing (AB'). The difference between the two paths is only small. However, the combined effect of countless electrons moving together in this way represents a significant net movement of charge.

EXTENSION

Variables that affect resistance

Effect of cross-sectional area and length on resistance

Understanding the way electrons move through a wire can help us make some predictions about the resistance of different objects.

For example, in a longer piece of wire, the electrons bump into more ions along the way, so more energy would be needed for the electrons to travel from one end to the other. In other words, a longer piece of wire would provide greater 'resistance' to the flow of electric current.

Similarly, a thicker piece of wire allows more electrons to flow through it at the same time, much like a dual-lane highway allows faster traffic flow than a single lane. In practice, the cross-sectional area of the wire (its area when viewed end on) is important. The greater the cross-sectional area of the wire, the lower its resistance will be.

Calculating the effect of length and area on resistance

The relationship between the resistance of a conductor and its length and thickness follows a mathematical relationship. There is a direct relationship between resistance and length: doubling the length of the conductor doubles its resistance. There is an inverse relationship between resistance and the cross-sectional area of the conductor. These relationships are captured in the equation:

$$R = \frac{\rho L}{A}$$

where R is resistance, L is length, A is cross-sectional area and ρ is resistivity, a property of the material from which the conductor is made.

Temperature and resistance

Another factor that affects the resistance of a material is its temperature. The temperature of an object is a measure of the average kinetic energy of its particles. The temperature of a solid is an indication of how quickly its particles are vibrating.

Increasing the temperature of a piece of copper wire means that the copper ions will vibrate back and forth more quickly. This makes it more likely that an electron will collide with the ion as it moves past it. Therefore, increasing the temperature of the wire also increases the resistance of the wire.

Similarly, current passing through a conductor can cause it to heat up. Think again of an electron moving through a copper wire: when the electron collides with a copper ion, it loses some of its kinetic energy. However, due to this collision, the copper ion gains kinetic energy, causing it to vibrate more quickly. An increase in the kinetic energy of the copper means that its temperature has increased, so the copper wire heats up.

This is one of the reasons why personal computers contain cooling fans, as shown in Figure 3.4.2. Electrical components are packed very tightly together on the computer motherboard. Cooling the components and the conductors that connect them prevents the computer from overheating. It also reduces the resistance of the components and helps them to run more efficiently.

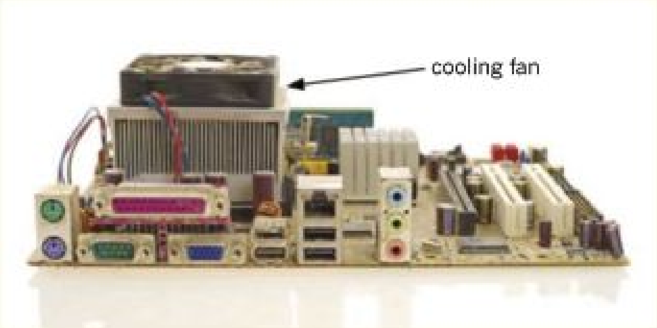


FIGURE 3.4.2 The cooling fan in this computer motherboard circulates air around the electrical components to cool them down.

Mathematically, the relationship is expressed as follows:

$$R = R_0[1+\alpha(T-T_0)]$$

where R is the resistance of the conductor at temperature T, R_0 is the resistance at temperature T_0 and α is the temperature coefficient of resistance (a property of each material which describes how big an impact changing temperature has on the resistance of the material). Temperature coefficient of resistance values are shown for some common conducting materials in Table 3.4.1.

Substance	copper	tungsten	nickel	iron	steel
$\alpha\,[\frac{1}{K}]$	3.9×10^{-3}	4.5×10^{-3}	6.0×10^{-3}	5.0×10^{-3}	3.0×10^{-3}

TABLE 3.4.1 Temperature coefficient of resistance values for some common conducting materials.

PHYSICS IN ACTION

Incandescent light bulbs

The complex relationship between electric current and temperature is put to use in a very common application: the incandescent light bulb (Figure 3.4.3).

FIGURE 3.4.3 An incandescent bulb produces light when its filament heats up.

An incandescent light bulb consists of a thin piece of curled or bent wire, called a *filament*, in a glass bulb. Often the bulb is evacuated (has the air removed) or filled with an inert (unreactive) gas so that it does not corrode. The wire is usually made of tungsten or another metal with a high melting point.

When an electric current passes through the filament, it heats up. This in turn increases the resistance of the filament, causing it to heat up further. The filament quickly becomes so hot that it starts to glow, radiating heat and light.

Traditionally, most household lighting was provided by incandescent light bulbs. However, this form of lighting is very inefficient. Only a small amount of the energy that goes into an incandescent light bulb is transformed into light: over 95% of the energy is lost as heat.

More recent inventions such as fluorescent tubes and LEDs (light-emitting diodes) are much more efficient and are increasingly being used as alternatives to incandescent light bulbs. Some examples of these are shown in Figure 3.4.4. In 2007, as a strategy to reduce carbon dioxide emissions, the Australian Government announced plans to phase out the use of incandescent bulbs.

FIGURE 3.4.4 Alternatives to incandescent light bulbs include fluorescent tubes, fluorescent bulbs and LED bulbs.

Many household electrical heating devices such as toasters and bar heaters work on a similar principle to the incandescent light bulb; although, in these situations, light is the unwanted or wasted energy.

OHM'S LAW

Georg Ohm (1789–1854) discovered that if the temperature of a metal wire was kept constant, the current flowing through it was directly proportional to the potential difference across it: mathematically, $I \propto V$. This relationship is known as Ohm's law. This relationship means that if the potential difference across a wire is doubled, for example, then the current flowing through the wire must also double. If the potential difference is tripled, then the current would also triple.

Ohm's law is usually written as:

ℹ $\Delta V = IR$ (or just $V = IR$)

where V is the potential difference in volts (V)

I is current in amps (A)

R is the constant of proportionality called resistance, in ohms (Ω).

This equation can be transposed to give a quantitative (mathematical) definition for resistance:

$$R = \frac{V}{I}$$

If an identical voltage produces two different sizes of current when separately connected to two light bulbs, then the resistance of the two light bulbs must differ. A higher current would mean a lower resistance of the light bulb, according to Ohm's law. This is because, when a conductor provides less resistance, more current can flow.

Worked example 3.4.1

USING OHM'S LAW TO CALCULATE RESISTANCE

When a potential difference of 3 V is applied across a piece of wire, 5 A of current flows through it. Calculate the resistance of the wire.

Thinking	Working
Ohm's law is used to calculate resistance.	$V = IR$
Rearrange the equation to find R.	$R = \frac{V}{I}$
Substitute in the values for this situation.	$R = \frac{3}{5}$ $= 0.6\ \Omega$

Worked example: Try yourself 3.4.1

USING OHM'S LAW TO CALCULATE RESISTANCE

An electric bar heater draws 10 A of current when connected to a 240 V power supply. Calculate the resistance of the element in the heater.

OHMIC AND NON-OHMIC CONDUCTORS

Conductors that obey Ohm's law are known as **ohmic** conductors. Ohmic conductors are usually called **resistors**.

An ohmic conductor can be identified by measuring the current that flows through the conductor when different potential differences are applied across it.

Worked example 3.4.2

USING OHM'S LAW TO CALCULATE RESISTANCE, CURRENT AND POTENTIAL DIFFERENCE

The table below shows measurements for the potential difference and corresponding current for an ohmic conductor.

V [V]	0	2	4	V_2
I [A]	0	0.25	I_1	0.75

Determine the missing results, I_1 and V_2.

Thinking	Working
Determine the factor by which potential difference has increased from the second column to the third column.	$\frac{4}{2} = 2$ The potential difference has doubled.
Apply the same factor increase to the current in the second column, to determine the current in the third column (I_1).	$I_1 = 2 \times 0.25$ $= 0.50$ A
Determine the factor by which current has increased from the second column to the fourth column.	$\frac{0.75}{0.25} = 3$ The current has tripled.
Apply the same factor increase to the potential difference in the second column, to determine the potential difference in the fourth column (V_2).	$V_2 = 3 \times 2$ $= 6$ V

Worked example: Try yourself 3.4.2

USING OHM'S LAW TO CALCULATE RESISTANCE, CURRENT AND POTENTIAL DIFFERENCE

The table below shows measurements for the potential difference and corresponding current for an ohmic conductor.

V [V]	0	3	9	V_2
I [A]	0	0.20	I_1	0.80

Determine the missing results, I_1 and V_2.

The data from an experiment in which the current and potential difference is measured for a device is usually plotted on an I–V graph. If the conductor is ohmic, this graph will be a straight line, as can be seen in Figure 3.4.5.

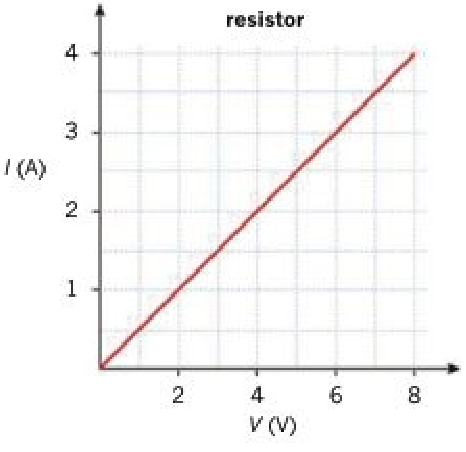


FIGURE 3.4.5 As the resistance of an ohmic conductor is constant, the I–V graph is a straight line.

The resistance of the ohmic conductor (or resistor) can be found from the gradient of the *I*–*V* graph. Ohm recognised that the gradient was equal to the inverse of the resistance:

$$\frac{1}{R} = \frac{\text{rise}}{\text{run}} = \frac{4-1}{8-2} = \frac{3}{6}$$

$$\therefore R = \frac{6}{3} = 2\ \Omega$$

However, not all conductors are ohmic. The *I*–*V* graphs for **non-ohmic** conductors are not straight lines (see Figure 3.4.6). Light bulbs and diodes are examples of non-ohmic conductors.

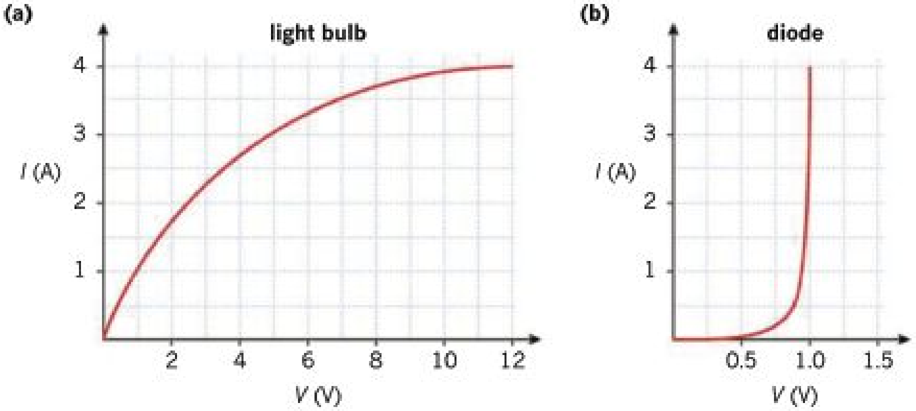


FIGURE 3.4.6 The *I*–*V* graph for a non-ohmic resistor is not a straight line.

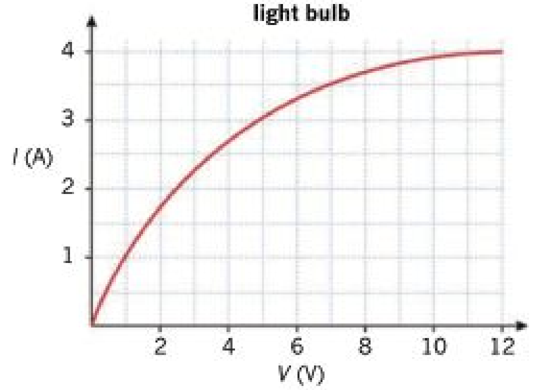
Using *I*–*V* graphs to determine resistance

The inverse of resistance is defined as the ratio *I*/*V*. For an ohmic conductor, this value will be a constant regardless of the potential difference across the conductor. However, the resistance of a non-ohmic conductor will vary. The resistance of a non-ohmic conductor for a particular potential difference can be found by determining the current flowing through the conductor at this value.

Worked example 3.4.3

CALCULATING RESISTANCE FOR A NON-OHMIC CONDUCTOR

Calculate the resistance of the light bulb with the *I*–*V* graph shown when the potential difference is 5.0 V.

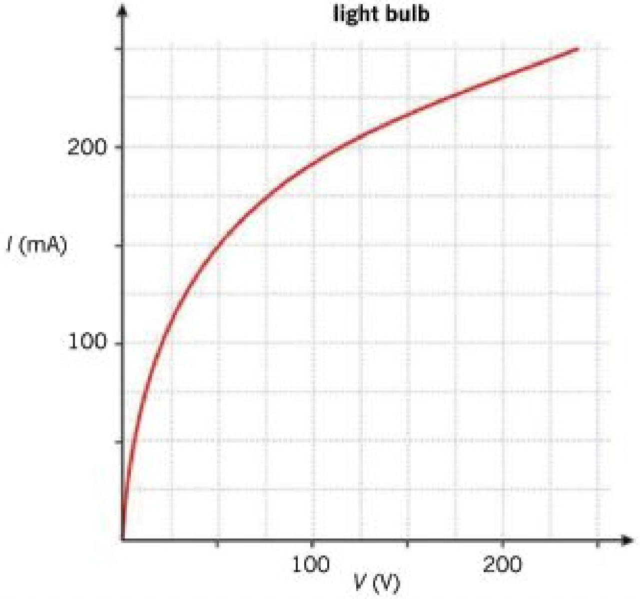


Thinking	Working
From the graph, determine the current at the required potential difference.	At $V = 5$ V, $I = 3$ A
Substitute these values into Ohm's law to find the resistance.	$R = \frac{V}{I} = \frac{5}{3} = 1.67\ \Omega$

Worked example: Try yourself 3.4.3

CALCULATING RESISTANCE FOR A NON-OHMIC CONDUCTOR

A 240 V, 60 W incandescent light bulb has *I–V* characteristics as shown in the graph. Calculate the resistance of the light bulb at 175 V.



RESISTORS IN SIMPLE CIRCUITS

Ohmic resistors are often used to control the amount of current in a particular circuit. Resistors can be manufactured to produce a relatively constant resistance over a range of temperatures. A colour-coding system is used on resistors to explain the amount of resistance they provide, including a percentage tolerance (precision). Figure 3.4.7 shows a resistor that uses the colour-coding system.

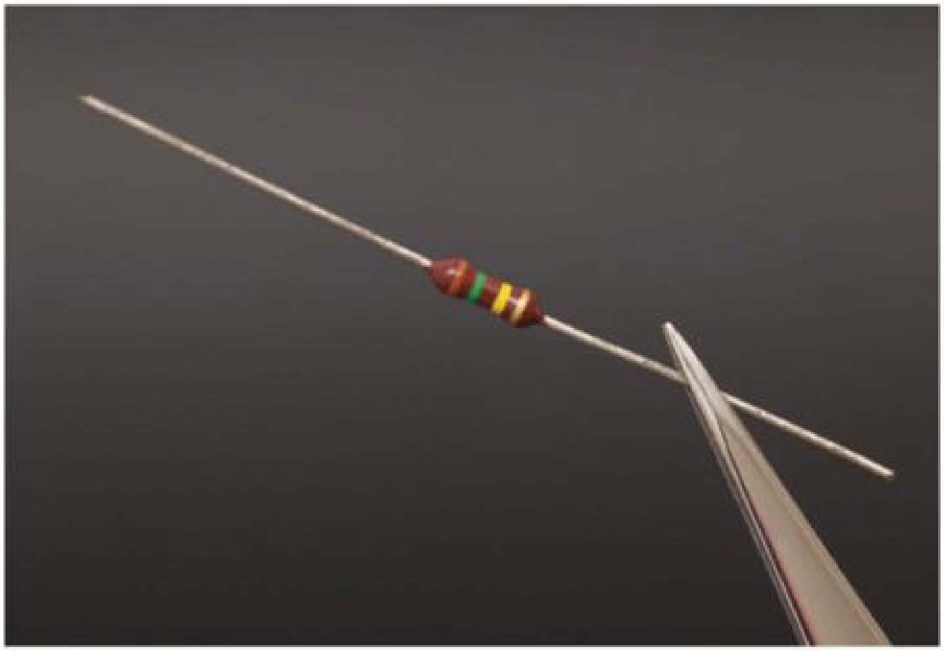


FIGURE 3.4.7 Common resistors are electrical devices with a known resistance. The coloured bands indicate the resistor's resistance and tolerance.

PHYSICSFILE

Colour-coded resistors

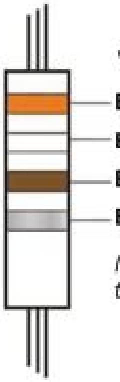
A resistor is typically a small piece of equipment which does not allow enough room to clearly print information about the resistor in the form of numbers. A colour-coding system is used on many resistors to convey detailed information in a small space about the resistance and tolerance of the resistor. Figure 3.4.8 below explains how to interpret the colour-coding system.

Resistor colour code

Band colour	Value
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Purple	7
Grey	8
White	9
Gold	0.1
Silver	0.01

Tolerance colour code

Band colour	±%
Brown	1
Red	2
Gold	5
Silver	10
None	20



What this means

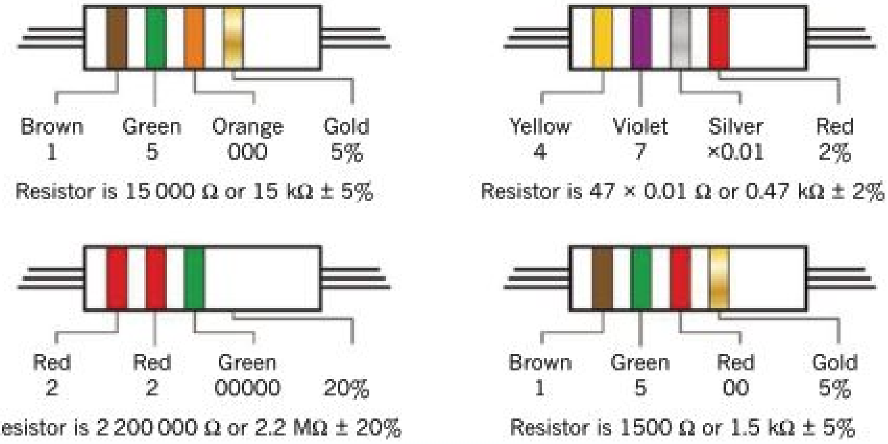
Band 1 First figure of value

Band 2 Second figure of value

Band 3 Number of zeros/multiplier

Band 4 Tolerance (±%) See below

Note that the bands are closer to one end than the other.



Resistor is 15 000 Ω or 15 kΩ ± 5%

Resistor is 47 × 0.01 Ω or 0.47 kΩ ± 2%

Resistor is 2 200 000 Ω or 2.2 MΩ ± 20%

Resistor is 1500 Ω or 1.5 kΩ ± 5%

FIGURE 3.4.8 Examples of resistor colour-coding.

Ohm's law can be used to determine the current flowing through a resistor when a particular potential difference is applied across it. Similarly, if the current and resistance are known, the potential difference across the resistor can be calculated.

Worked example 3.4.4

USING OHM'S LAW TO FIND CURRENT

A 100 Ω resistor is connected to a 12 V battery. Calculate the current (in mA) that would flow through the resistor.

Thinking	Working
Recall Ohm's law.	$V = IR$
Rearrange the equation to make I the subject.	$I = \frac{V}{R}$
Substitute in the values for this problem and solve.	$I = \frac{12}{100} = 0.12$ A
Convert the answer to the required units.	$I = 0.12$ A $= 0.12 \times 10^3$ mA $= 120$ mA

Worked example: Try yourself 3.4.4

USING OHM'S LAW TO FIND CURRENT

The element of a bar heater has a resistance of 25 Ω. Calculate the current (in mA) that would flow through this element if it is connected to a 240 V supply.

Worked example 3.4.5

USING OHM'S LAW TO FIND POTENTIAL DIFFERENCE

A current of 0.25 A flows through a 22 Ω resistor. Calculate the voltage across the resistor. Give your answer correct to one decimal place.

Thinking	Working
Recall Ohm's law.	$V = IR$
Substitute in the values for this problem and solve.	$V = 0.25 \times 22$ $= 5.5$ V

Worked example: Try yourself 3.4.5

USING OHM'S LAW TO FIND POTENTIAL DIFFERENCE

The globe of a torch has a resistance of 5.7 Ω when it draws 700 mA of current. Calculate the potential difference across the globe.

3.4 Review

SUMMARY

- Resistance is a measure of how hard it is for current to flow through a particular material. Resistance is measured in ohms (Ω).
- The resistance of a material depends on its length, cross-sectional area and temperature.
- Ohm's law describes the relationship between current, potential difference and resistance: $V = IR$
- Ohmic conductors have a constant resistance. The resistance of non-ohmic conductors varies for different potential differences.

KEY QUESTIONS

1 An experiment is conducted to gather data about the relationship between current and potential difference for three ohmic devices, labelled A, B and C. The data is used to plot an I–V graph for each device, as shown in Figure 3.4.9.

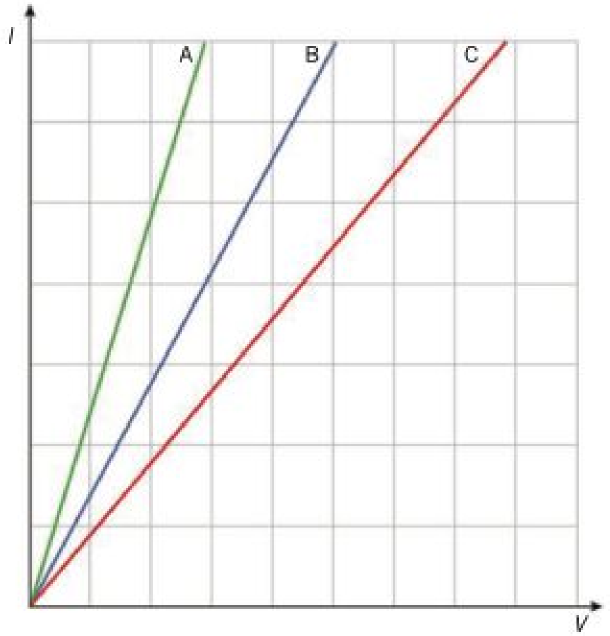


FIGURE 3.4.9

a For a given potential difference, list the devices in order of highest current to lowest current.

b List the devices in order of highest resistance to lowest resistance.

2 Table 3.4.2 below shows measurements for the potential difference and corresponding current for an ohmic conductor.

V [V]	0	2	3	V_2
I [A]	0	0.25	I_1	0.60

TABLE 3.4.2

Determine the missing results, I_1 and V_2.

3 A student obtains a graph of the current–voltage characteristics of a piece of resistance wire (see Figure 3.4.10).

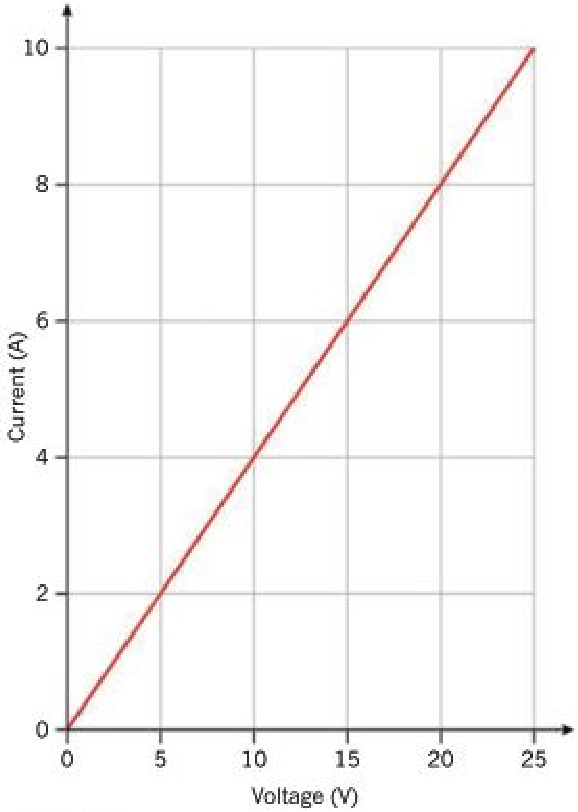


FIGURE 3.4.10

a Explain whether this piece of wire is ohmic or non-ohmic.

b What current flows in this wire at a voltage of 7.5 V?

c What is the resistance of this wire?

3.4 Review *continued*

4 A student finds that the current through a resistor is 3.5 A when a voltage of 2.5 V is applied to it.

- **a** What is the resistance?
- **b** The voltage is then doubled and the current is found to increase to 7.0 A. Is the resistor ohmic or not?

5 Rose and Rachel are trying to find the resistance of an electrical device. They find that at 5 V it draws a current of 200 mA and at 10 V it draws a current of 500 mA. Rose says that the resistance is 25 Ω, but Rachel maintains that it is 20 Ω. Who is right and why?

6 Nick has an ohmic resistor to which he has applied 5 V. He measures the current as 45 mA. He then increases the voltage to 8 V. What current will he find now?

7 Lisa finds that when she increases the voltage across an ohmic resistor from 6 V to 10 V the current increases by 2 A.

- **a** What is the resistance of this resistor?
- **b** What current does it draw at 10 V?

8 The resistance of a piece of wire is found to be 0.8 Ω. What would be the resistance of:

- **a** a piece of the same wire twice as long?
- **b** a piece of wire of twice the diameter?

9 A strange electrical device has the *I–V* characteristics shown in Figure 3.4.11.

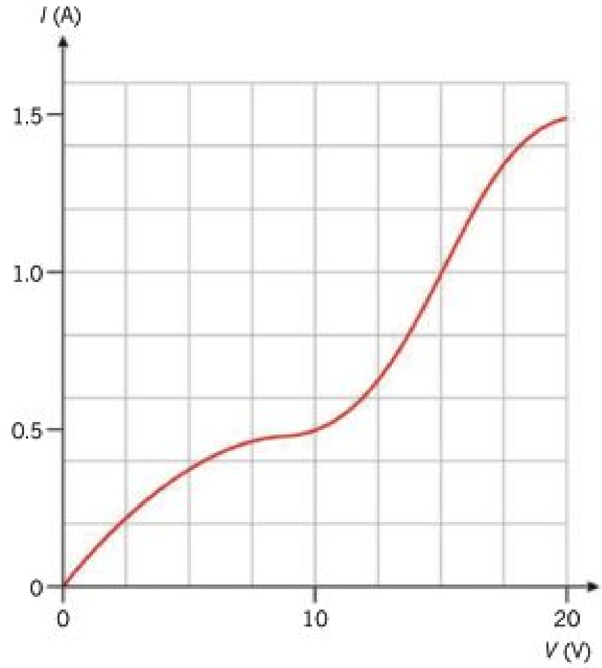


FIGURE 3.4.11

- **a** Is it an ohmic or non-ohmic device? Explain.
- **b** What current is drawn when a voltage of 10 V is applied to it?
- **c** What voltage would be required to double the current drawn at 10 V?
- **d** What is the resistance of the device at:
  - **i** 10 V?
  - **ii** 20 V?

Chapter review

03

KEY TERMS

ammeter
charge
conductor
conventional current
coulomb
current
electric circuit
electrical potential energy
electricity
electron flow
elementary charge
insulators
ion
ionised
metals
net charge
non-metals
non-ohmic
ohmic
parallel circuit
potential difference
power
resistance
resistor
transfer
transform
voltmeter
volts

1 Approximately how many electrons make up a charge of -3 C?

2 What will be the approximate charge on 4.2×10^{19} protons?

3 Which charged particles are moving when an electric current flows in a circuit?
- **A** negatively charged electrons
- **B** positively charged electrons
- **C** positively charged protons
- **D** both negative and positive charges

4 An alpha particle consists of two protons and two neutrons. Calculate the charge on an alpha particle.

5 Why is water flowing in a pipe a common analogy for electric current?
- **A** Water does not conduct electricity.
- **B** Water can leak out of a pipe.
- **C** Water is not compressed or lost as it flows through a pipe.
- **D** Water and electricity do not mix.

6 What does the bicycle chain analogy show?
- **A** Electrons physically touch one another.
- **B** Only a small number of electrons move.
- **C** Electrons are the same shape as the links in a bicycle chain.
- **D** Electrons move simultaneously in every part of the circuit.

7 Calculate the current that flows when 0.23 C of charge passes a point in a circuit each minute.

8 Compare the meaning of the terms 'conventional current' and 'electron flow'.

9 A current of 1.6 A flows for 100 seconds. Calculate:
- **a** the amount of charge, in coulombs, that moves past a point in this time
- **b** the number of electrons that move past a point in this time.

10 A current of 0.04 A flows for a certain amount of time. In this time 5×10^{18} electrons move past a point. Calculate:
- **a** the amount of charge, in coulombs, that moves past a point
- **b** the amount of time that the current is flowing.

11 A phone battery has a voltage of 3.8 V. If 2 C of charge is drawn from the battery, what amount of energy would this provide?

12 A battery does 2 joules of work on a charge of 0.5 coulombs to move it from point A to point B. Calculate the potential difference between the two points A and B.

13 Which quantities would you need to measure to calculate the amount of electrical energy used to heat water using an electric element?
- **A** potential difference, resistance and current
- **B** time, current and charge
- **C** current, time and potential difference
- **D** potential difference and current

14 How much power does an appliance use if it does 2500 J of work in 30 minutes?

15 A battery gives a single electron 1.4×10^{-18} J of energy. Calculate the potential difference supplied by the battery to two decimal places.

16 A 230 V appliance consumes 2000 W of power. The appliance is left on for 2 hours. What current flows through the appliance?

17 A student finds that the current through a wire is 5 A while a voltage of 2.5 V is applied to it. Calculate the resistance of the wire.

18 A 60 W incandescent globe draws 0.25 A when connected to a 240 V power supply. Calculate the resistance of the globe.

Chapter review *continued*

19 Calculate the resistance at 50 V of the non-ohmic conductor with the *I–V* graph shown in Figure 3.5.1.

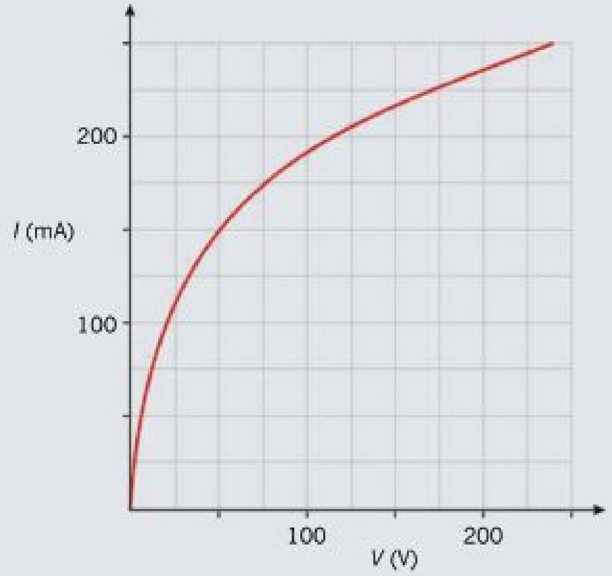


FIGURE 3.5.1

20 A current of 0.25 A flows through an 80 Ω resistor. Calculate the voltage across it.

21 When 1.5 V is applied across a particular resistor, the current through the resistor is 50 mA. What is the resistance of the resistor?

22 Calculate the resistance of the non-ohmic conductor with the *I–V* graph shown in Figure 3.5.2 at the following voltages:

a 1.0 V
b 7.0 V
c 12.0 V.

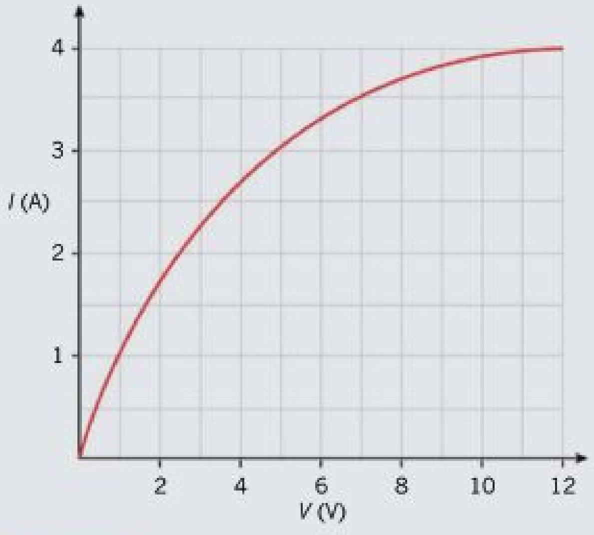


FIGURE 3.5.2

23 Explain why even a good conductor such as copper wire provides some resistance to current.

24 A potential difference of 240 V is used to generate a current of 5 A to heat water for 3 minutes. Calculate the energy transferred to the water in that time.

25 Siobhan is measuring the current flowing through a non-ohmic conductor as she varies the potential difference of the power source. The results are shown in the graph in Figure 3.5.3. Calculate the resistance when the potential difference is 1.0 V.

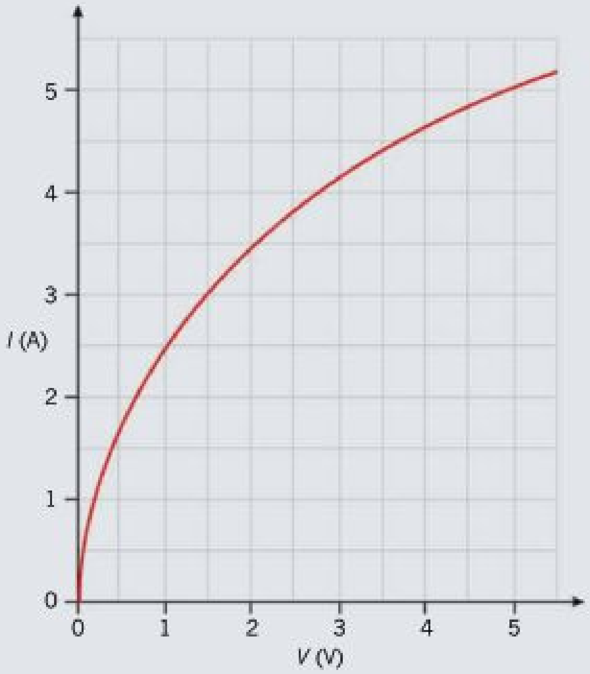


FIGURE 3.5.3

26 A laptop computer draws 500 mA from a 240 V power point. What is the power output of the computer?

27 A hair dryer is designed to produce 1600 W of power when connected to a 240 V power supply. What is the resistance of the hair dryer?

28 A 120 V lamp draws 12.5 A when cold but only 0.8 A when hot. Calculate the resistance of the lamp at each temperature.

29 If 0.8 A of current flows through a light bulb, calculate how many electrons enter the light bulb each second.

30 A 3 V torch with a 0.3 A bulb is switched on for 1 minute.

a How much charge has travelled through the filament in this time?
b How much energy has been used?
c Where has this energy come from?

CHAPTER

04 Practical electric circuits

Electric circuits are the basis of much of our modern society. This chapter introduces a range of circuits, from simple series circuits to the complex parallel wiring systems that make up a modern home. Electric circuits can be used to perform energy transfers and transformations through devices such as light bulbs, thermistors, light-dependent resistors and light-emitting diodes. It is essential that anyone working with electricity learns how to do so safely in the home and laboratory. Several safety mechanisms are examined that minimise the effect of current on humans.

Key knowledge

By the end of this chapter, you will have covered material from the study of practical electric circuits, and will be able to:

- apply the kilowatt-hour (kWh) as a unit of energy
- model resistance in series and parallel circuits using equivalent effective resistance in arrangements in
  - series: $R_T = R_1 + R_2 + \cdots + R_n$
  - parallel: $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}$
- calculate and analyse the effective resistance of circuits comprising parallel and series resistance and voltage dividers
- model household (AC) electrical systems as simple direct current (DC) circuits
- compare power transfers in series and parallel circuits
- explain why the circuits in homes are mostly parallel circuits
- investigate and apply theoretically and practically concepts of current, resistance, potential difference (voltage drop) and power to the operation of electronic circuits comprising resistors, light bulbs, diodes, thermistors, light-dependent resistors (LDRs), light-emitting diodes (LEDs) and potentiometers
- investigate practically the operation of simple circuits containing resistors, variable resistors, diodes and other non-ohmic devices
- describe energy transfers and transformations in reference to transducers
- model household electricity connections as a simple circuit comprising fuses, switches, circuit breakers, loads and earth
- compare the operation of safety devices including fuses, circuit breakers and residual-current devices (RCDs)
- describe the causes, effects and treatment of electric shock in homes and identify the approximate danger thresholds for current and duration.

4.1 Series and parallel circuits

When a circuit contains more than one resistor, Ohm's law alone is not sufficient to predict the current flowing through and the potential difference across each resistor. Additional concepts such as Kirchhoff's rules and the idea of equivalent resistance can be used to analyse these complex, multi-component circuits.

No matter how complex a circuit, it can always be broken up into sections in which the circuit elements are in series or parallel. This section investigates the difference between these two types of circuits.

RESISTORS IN SERIES

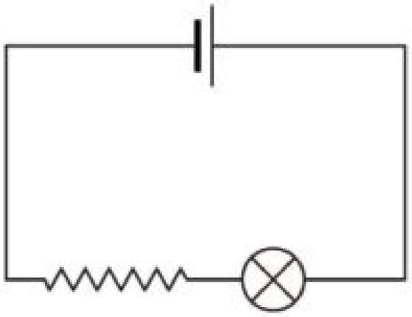


FIGURE 4.1.1 This circuit has a resistor and light bulb connected in series.

Some circuits contain more than one electrical component. When these components are connected one after another in a continuous loop, this is called a **series circuit**. Components connected in this way are said to have been connected 'in series'. The circuit shown in Figure 4.1.1 shows a resistor and a light bulb connected in series with an electric cell.

Series circuits are very easy to construct, but they have some disadvantages. As every component is connected one after the other, each component is dependent on each other. If one component is removed or breaks down, the circuit is no longer a closed loop and it won't work. Figure 4.1.2 shows how removing a globe from a series circuit interrupts the entire circuit. Due to this characteristic, series circuits with more than one component are not commonly used in the home.

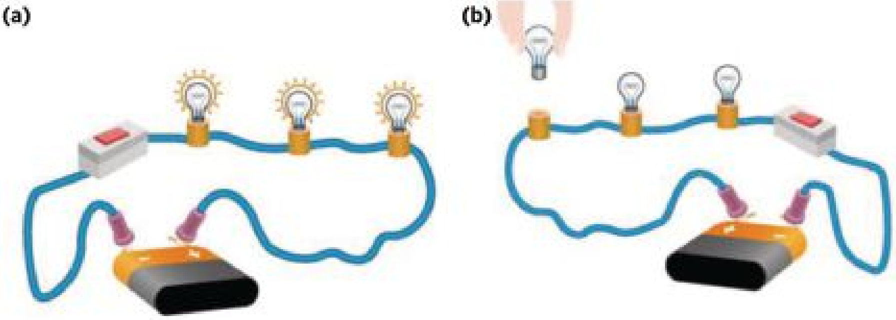


FIGURE 4.1.2 (a) All components in the circuit are intact and so the circuit is a closed loop. (b) When one light bulb is removed, the whole circuit is interrupted.

Conservation of charge

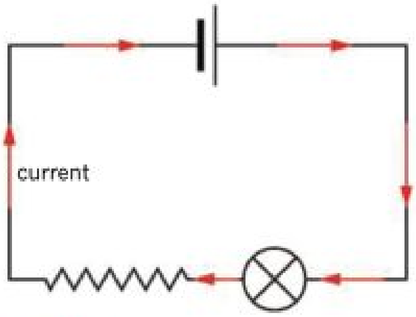


FIGURE 4.1.3 In a series circuit, the same current flows through each component.

When analysing a series circuit it is important to understand that the same amount of current flows in every part of the circuit. Since electric charges are not created or destroyed within an electric circuit, the current flowing out of the cell must be the same as the current flowing through the lamp, which is also the same as the current flowing through the resistor. This current also flows unchanged back into the cell as shown in Figure 4.1.3.

The current in a series circuit is the same in every part of the circuit.

Remember that, by convention, the current is represented as flowing from the positive terminal of the cell to the negative terminal. Electrons move in the opposite direction.

Kirchhoff's loop rule

Kirchhoff's loop rule says that the sum of the potential differences across all the elements around any closed circuit loop must be zero. This means that the total potential drop around a closed circuit must be equal to the total potential gain in the circuit. For example, if a battery provides 9 V to a circuit, then the sum of all of the potential drops across the components must add to 9 V.

This rule is essentially another version of the law of conservation of energy.

The energy given to the charges (potential gain) must be equal to the energy lost by the charges (potential drop). In a series circuit, the energy loss will be spread across a number of different components.

Figure 4.1.4 shows how the voltage provided by the battery is shared across a resistor and a lamp. The power supply in the figure is labelled EMF. Devices that are a source of energy for a circuit are referred to as sources of EMF or electromotive force. EMF, measured in volts (V), is another term for the work done on charges to provide a potential difference between the terminals of the power supply.

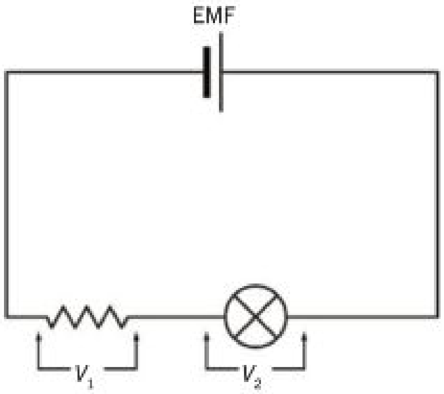


FIGURE 4.1.4 Kirchhoff's loop rule. In this series circuit, the sum of the potential drops across the resistor and the lamp (i.e. $V_1 + V_2$) will be equal to the potential difference (EMF) provided by the battery.

There are a number of ways to visualise the energy changes in this circuit. One common analogy is to think of the charges as water being pumped around an elevated water course. The water gains potential energy as it is pumped higher, and as it flows back down the potential energy is converted into other forms. The diagram in Figure 4.1.5 shows how the analogy works with the energy changes that occur in a circuit. The battery acts as a 'pump' that pushes electrons up to a higher energy level and the electrons gain potential energy. As the electrons pass down through components in the circuit, their energy is transformed into other forms.

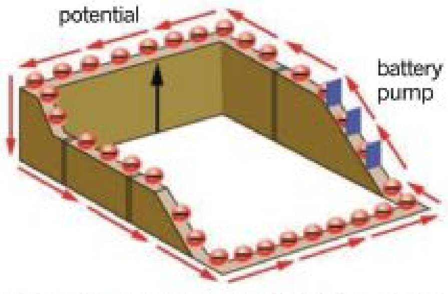


FIGURE 4.1.5 An analogy for analysing a circuit: the battery acts as a 'pump' which transfers potential energy to electrons. The electrons lose potential energy as they flow 'down' through components in the circuit.

The change in electrical energy available to electrons can also be represented graphically, as shown in Figure 4.1.6.

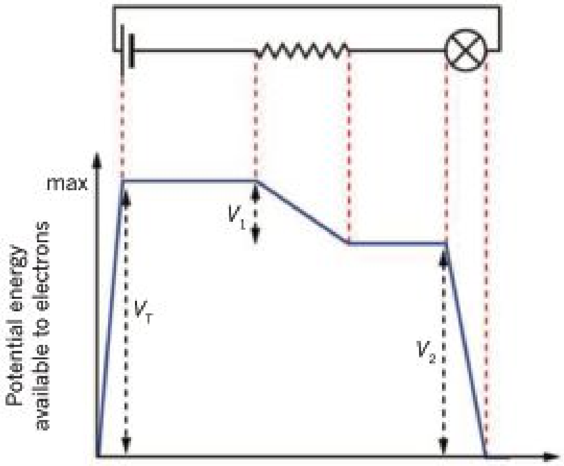


FIGURE 4.1.6 The electric potential energy of an electron changes as it moves around the circuit. Some of this energy is lost as the electrons pass through the resistor. The remaining energy is lost as the electrons pass through the bulb. In this circuit, the bulb has more resistance than the resistor.

Equivalent series resistance

Consider the circuit in Figure 4.1.7. If the resistance of the fixed resistor is R_1, the resistance of the lamp is R_2 and the current flowing through both of them is I, then Ohm's law gives:

$V_1 = I \times R_1$

and

$V_2 = I \times R_2$

The total voltage drop across the two components is:

$V_{\text{Total}} = V_1 + V_2 = IR_1 + IR_2 = I \times (R_1 + R_2)$

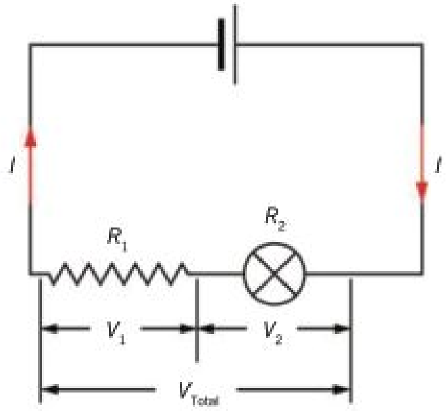


FIGURE 4.1.7 Using Ohm's law, it is possible to show the relationship between the potential difference supplied by the cell, V_{Total}, the current flowing in the circuit, I, and the resistances of the two components R_1 and R_2.

This equation shows the relationship between the potential difference supplied by the cell and the potential differences of the lamp and resistor. The last part of the equation also shows that the lamp and resistor can be replaced with a single resistor, without changing the current in the circuit. The single resistor needs to have a total resistance of $R_1 + R_2$.

In general, a number of individual resistors connected in series can be replaced by an equivalent **effective resistance** (also called the total resistance, R_T) equal to the sum of the individual resistances. Figure 4.1.8 on page 126 shows how two resistors can be replaced with a single one.

$R_T = R_1 + R_2 + \ldots + R_n$

where R_T is the equivalent effective series resistance and R_1, R_2,…, R_n are the individual resistances.

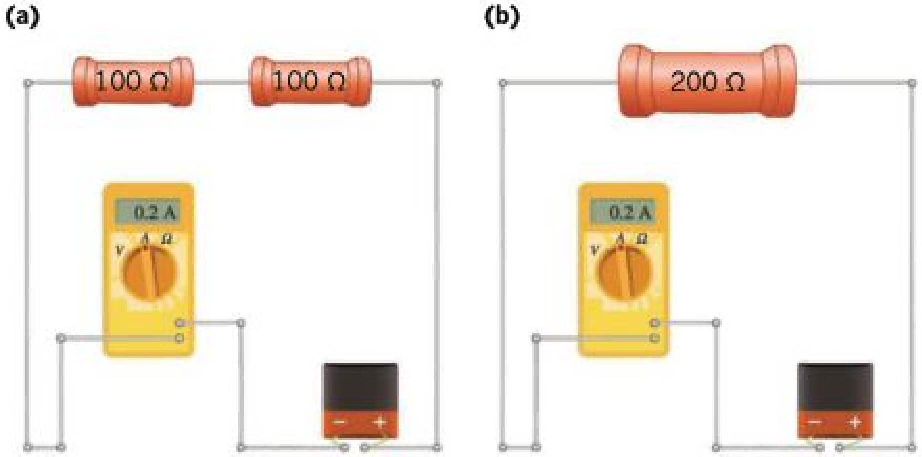


FIGURE 4.1.8 Two 100 Ω resistors (a) can be replaced with a single 200 Ω resistor (b) to have the same effect in a circuit.

Equivalent resistances can be used in circuit analysis to simplify a complicated circuit diagram so that current and potential difference can be determined.

Worked example 4.1.1

CALCULATING AN EQUIVALENT SERIES RESISTANCE

A 100 Ω resistor is connected in series with a 690 Ω resistor and a 1.2 kΩ resistor. Calculate the equivalent series resistance.

Thinking	Working
Recall the formula for equivalent series resistance.	$R_T = R_1 + R_2 + \ldots + R_n$
Substitute in the given values for resistance. Make sure to convert kΩ to Ω. Solve to find the equivalent series resistance.	$R_T = 100 + 690 + 1200 = 1990\ \Omega$

Worked example: Try yourself 4.1.1

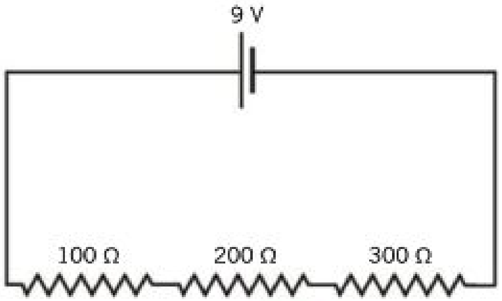
CALCULATING AN EQUIVALENT SERIES RESISTANCE

A string of Christmas lights consists of 20 light bulbs connected in series. Each bulb has a resistance of 8 Ω. Calculate the equivalent series resistance of the Christmas lights.

Worked example 4.1.2

USING EQUIVALENT SERIES RESISTANCE FOR CIRCUIT ANALYSIS

Use an equivalent series resistance to calculate the current flowing in the series circuit below and the potential difference across each resistor.

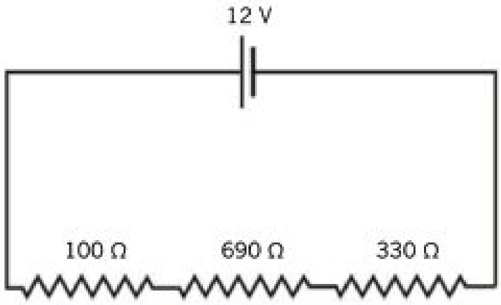


Thinking	**Working**
Recall the formula for equivalent series resistance.	$R_T = R_1 + R_2 + R_3 + ... + R_n$
Find the equivalent (total) resistance in the circuit.	$R_T = 100 + 200 + 300 = 600\ \Omega$
Use Ohm's law to calculate the current in the circuit. Whenever calculating current in a series circuit, use R_T and the voltage of the power supply.	$I = \frac{V}{R}$ $= \frac{9}{600}$ $= 0.015$ A
Use Ohm's law to calculate the potential difference across each separate resistor.	$V = IR$ so $V_1 = 0.015\ \text{A} \times 100\ \Omega = 1.5\ \text{V}$ $V_2 = 0.015\ \text{A} \times 200\ \Omega = 3.0\ \text{V}$ $V_3 = 0.015\ \text{A} \times 300\ \Omega = 4.5\ \text{V}$
Use the loop rule to check the answer.	$V_T = V_1 + V_2 + V_3$ $= 1.5\ \text{V} + 3.0\ \text{V} + 4.5\ \text{V}$ $= 9.0\ \text{V}$ Since this is the same as the voltage provided by the cell, the answer is reasonable.

Worked example: Try yourself 4.1.2

USING EQUIVALENT SERIES RESISTANCE FOR CIRCUIT ANALYSIS

Use an equivalent series resistance to calculate the current flowing in the series circuit below and the potential difference across each resistor.



RESISTORS IN PARALLEL

One of the disadvantages of series circuits is that if a switch is opened or a device disconnected, then the circuit is broken and current stops flowing. In everyday life, we often want to switch devices on and off independently. **Parallel circuits** allow us to do this.

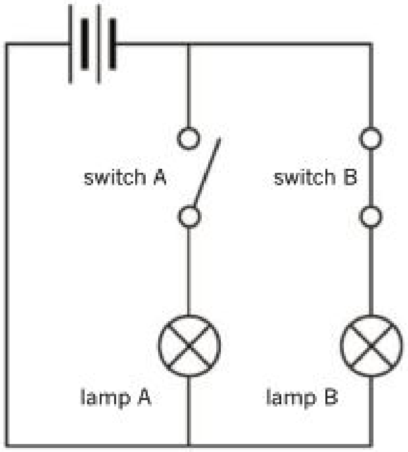


FIGURE 4.1.9 In this parallel circuit, lamp A will be off and lamp B will be on.

The circuit diagram in Figure 4.1.9 shows a simple parallel circuit. Even if switch A is open (as shown), lamp B will still light up as it is part of an unbroken complete circuit including the battery. Similarly, if switch A is closed and switch B is opened, current will light up lamp A and not lamp B. Alternatively, both switches could be closed to light up both lamps or both switches could be opened to switch both lamps off.

In a series circuit, all the components are in the same loop and therefore the same current flows through each component. In comparison, each loop of a parallel circuit acts like an independent circuit with its own current.

Consider again the water analogy. When water flows through pipes, it is not lost. If the pipe splits in two, some of the water flows in one pipe, and the remaining water will flow in the other pipe. If the two sections re-join, the water comes back together again, just as it was before the pipe was split. The same occurs with charges flowing in a parallel circuit. Figure 4.1.10 shows how the charges go through one globe or the other. This means that while the current in the main part of the circuit remains constant, in the parallel section, the current is divided between each branch. The readings on both ammeters, A_1 and A_2, will be the same.

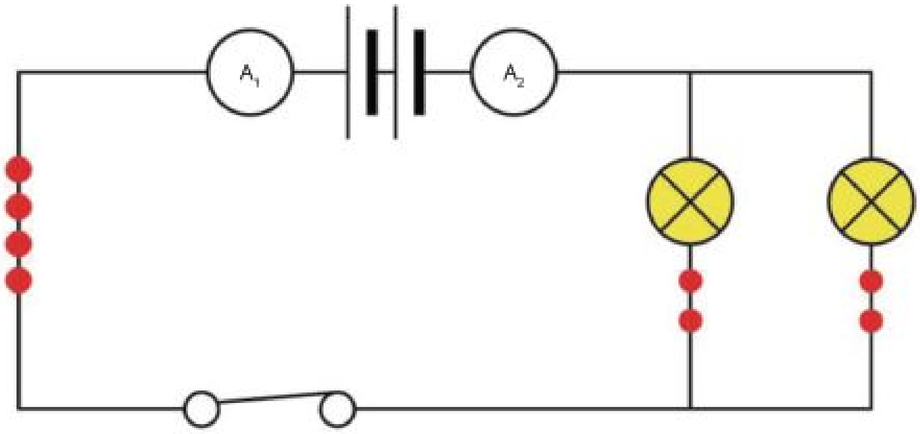


FIGURE 4.1.10 Unlike a series circuit, charges flowing in a parallel circuit have a choice of path.

The current in the main part of a parallel circuit is the sum of the currents in each branch of the circuit.

$I_T = I_1 + I_2 + \ldots + I_n$

Unlike a series circuit, in a parallel circuit, the voltage is not shared between resistors; the voltage is the same across each branch. This is because, while the charges take different pathways, they have the same amount of energy no matter which path they take. An advantage of parallel circuits is that globes connected in this way are brighter than if they were connected in series. The potential energy of the charges is not shared between the globes.

The voltage is the same in each branch of a parallel circuit.

Kirchhoff's junction rule

Parallel circuits involve **junctions** where current can flow in a variety of directions. The behaviour of current at these points is predicted by Kirchhoff's junction rule:

> ℹ The total amount of current flowing into a junction must be the same as the total current flowing out of the junction.

This rule is just an extension of the idea of conservation of charge; that is, that charges cannot be created or destroyed. Although the number of electrons flowing into a junction might be very large, electrons are not created or destroyed in the junction so the same number of electrons must flow out again. This is illustrated in Figure 4.1.11. Kirchhoff's junction rule explains how current splits in a parallel circuit. It explains why the current in the main part of the circuit is the sum of the currents in each parallel branch.

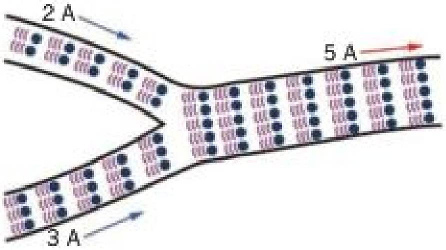


FIGURE 4.1.11 The current flowing into any junction must be equal to the current flowing out of it.

Equivalent parallel resistance

When additional resistors are added in a series circuit, the total resistance of the circuit increases. Additional resistance means that less current flows through the circuit.

In contrast, adding an additional resistor in parallel means that more current flows through the circuit because another path for charges has been added. This means that the total resistance for the circuit decreases.

Parallel circuits are more complicated than series circuits, hence the formula used to calculate the equivalent effective (total) resistance is more complicated.

> ℹ $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$
>
> where R_T is the equivalent effective resistance and $R_1, R_2, \ldots R_n$ are the individual resistances.

Worked example 4.1.3

CALCULATING AN EQUIVALENT PARALLEL RESISTANCE

A 100 Ω resistor is connected in parallel with a 300 Ω resistor. Calculate the equivalent parallel resistance.	
Thinking	**Working**
Recall the formula for equivalent effective resistance.	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$
Substitute in the given values for resistance.	$\frac{1}{R_T} = \frac{1}{100} + \frac{1}{300}$
Solve for R_T.	$\frac{1}{R_T} = \frac{1}{100} + \frac{1}{300}$ $= \frac{3}{300} + \frac{1}{300}$ $= \frac{4}{300}$ $R_T = \frac{300}{4}$ $= 75\ \Omega$

Worked example: Try yourself 4.1.3

CALCULATING AN EQUIVALENT PARALLEL RESISTANCE

A 20 Ω resistor is connected in parallel with a 50 Ω resistor. Calculate the equivalent parallel resistance.

PHYSICSFILE

Kirchhoff's contributions

Both the junction rule discussed here and the loop rule described earlier were first discovered by the German physicist Gustav Kirchhoff (1824–87). Kirchhoff also made important contributions in the fields of spectroscopy, thermochemistry and the study of black-body radiation. He worked with Robert Bunsen, the German chemist who developed the Bunsen burner.

FIGURE 4.1.12 Gustav Kirchhoff discovered the rules that underpin our understanding of how electric circuits work.

The effective (total) resistance of a set of resistors connected in parallel will always be smaller than the smallest resistor in the set.

In a parallel circuit:
$R_{\text{Total}} < R_{\text{smallest resistor}}$

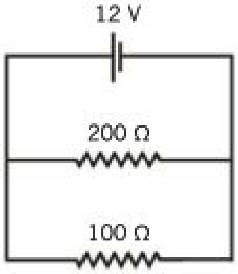
Notice that in the previous worked example, the equivalent effective resistance was smaller than the smallest individual resistance. This is because adding a resistor provides an additional pathway for current. Since more current flows, the resistance of the circuit has been effectively reduced.

If you consider the smallest resistor in any parallel combination, say, the 20 Ω resistor in Worked example: Try yourself 4.1.3 on page 129, the addition of the 50 Ω resistor in parallel with it allows the current an extra pathway and therefore it is easier for the current to flow through the combination. The effective resistance of the pair must be less than the 20 Ω alone.

Worked example 4.1.4

USING EQUIVALENT PARALLEL RESISTANCE FOR CIRCUIT ANALYSIS

Find an equivalent parallel resistance to calculate the current flowing out of the 12 V cell in the parallel circuit shown. Also find the current flowing through each resistor.

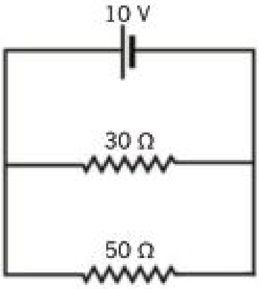


Thinking	**Working**
Recall the formula for equivalent parallel resistance.	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$
Substitute in the given values for resistance.	$\frac{1}{R_T} = \frac{1}{100} + \frac{1}{200}$
Solve for R_T.	$\frac{1}{R_T} = \frac{1}{100} + \frac{1}{200}$ $= \frac{2}{200} + \frac{1}{200}$ $= \frac{3}{200}$ $R_T = \frac{200}{3}$ $= 67\ \Omega$
Use Ohm's law to calculate the current in the circuit. To calculate *I*, use the voltage of the power supply and the total resistance.	$I_{\text{circuit}} = \frac{V}{R} = \frac{12}{67} = 0.18$ A
Use Ohm's law to calculate the current through each separate resistor. Remember that the voltage through each resistor is the same as the voltage of the power supply, 12 V in this case.	100 Ω resistor: $I_{100} = \frac{V}{R} = \frac{12}{100} = 0.12$ A 200 Ω resistor: $I_{200} = \frac{V}{R} = \frac{12}{200} = 0.060$ A
Use the junction rule to check the answers.	$I_{\text{circuit}} = I_{100} + I_{200}$ 0.18 A = 0.12 A + 0.060 A This is correct, so the answers are reasonable.

Worked example: Try yourself 4.1.4

USING EQUIVALENT PARALLEL RESISTANCE FOR CIRCUIT ANALYSIS

Use an equivalent parallel resistance to calculate the current flowing in the parallel circuit below and through each resistor of the circuit.



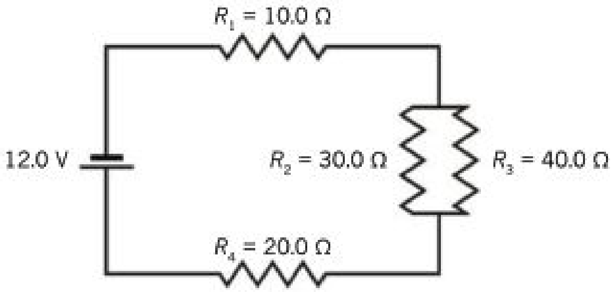
COMPLEX CIRCUIT ANALYSIS

Some circuits combine elements of series wiring and parallel wiring. A general strategy for analysing these circuits is to reduce the complex circuit to a single equivalent resistance to determine the current drawn by the circuit. It is then possible to step back through the process of simplification to analyse each section of the circuit as needed.

Worked example 4.1.5

COMPLEX CIRCUIT ANALYSIS

Calculate the potential difference across and current through each resistor in the circuit below.



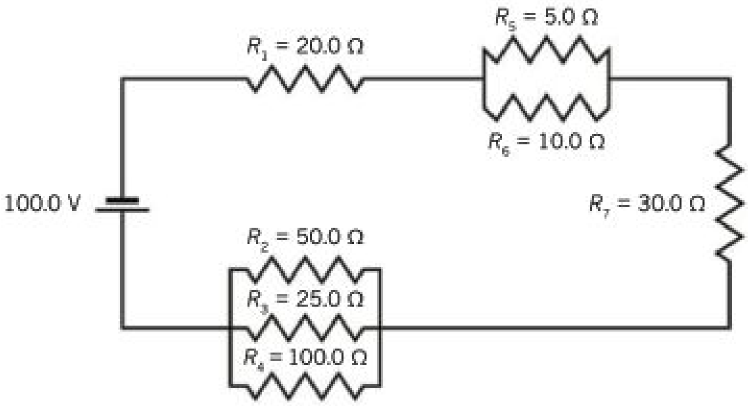
Thinking	Working
Find an equivalent resistance for the parallel resistors. The effective resistance of these should be less than the smaller resistor, that is, smaller than 30 Ω.	$\frac{1}{R_{2-3}} = \frac{1}{R_2} + \frac{1}{R_3}$ $= \frac{1}{30.0} + \frac{1}{40.0}$ $= \frac{4}{120.0} + \frac{3}{120.0}$ $R_{2-3} = \frac{120.0}{7} = 17.1\ \Omega$
Find an equivalent series resistance for the circuit as the circuit can now be thought of as three resistors in series: 10.0 Ω, 17.1 Ω and 20.0 Ω.	$R_T = 10.0\ \Omega + 17.1\ \Omega + 20.0\ \Omega$ $= 47.1\ \Omega$
Use Ohm's law to calculate the current in the circuit. Use the supply voltage and total resistance to do this calculation.	$V = IR$ $I = \frac{V}{R} = \frac{12.0}{47.1}$ $= 0.255\ \text{A}$

Use Ohm's law to calculate the potential difference across each resistor (or parallel group of resistors) in series. (Note that the potential difference across R_2 is the same as that across R_3 as they are in parallel.)	$V = IR$ $V_1 = 0.255 \times 10.0 = 2.55$ V $V_{2\text{-}3} = 0.255 \times 17.1 = 4.36$ V $V_4 = 0.255 \times 20.0 = 5.10$ V Check: $2.55 + 4.36 + 5.10 \approx 12.0$ V (with some slight rounding error) This confirms that the loop rule holds for this circuit.
Use Ohm's law where necessary to calculate the current through each resistor.	$I_1 = I_4 = 0.255$ A $I = \frac{V}{R}$ $I_2 = \frac{4.36}{30.0} = 0.145$ A $I_3 = \frac{4.36}{40.0} = 0.109$ A Check: $0.145 + 0.109 \approx 0.255$ A (with some slight rounding error) This confirms that the junction rule holds for this section.

Worked example: Try yourself 4.1.5

COMPLEX CIRCUIT ANALYSIS

Calculate the potential difference across and the current through each resistor in the circuit below.



PHYSICSFILE

Identical resistors in parallel

Where identical resistors are placed in parallel, the total resistance of the combination can be found by simply dividing the value of one of the resistors by the number of resistors.

For example, the total resistance of three 12 Ω resistors connected in parallel would have an effective resistance of 4 Ω.

$R_T = 12 \div 3 = 4\ \Omega$.

The three 12 Ω resistors in parallel could be replaced with a single 4 Ω resistor.

Similarly, the equivalent resistance of two 10 Ω resistors placed in parallel would be 5 Ω.

PHYSICS IN ACTION

Superconductors

In 1908 Dutch physicist Kamerlingh Onnes (1853–1926) was the first to liquefy helium. This occurred at 4.2 K (–269°C). Onnes was the first to liquefy a number of gases. He was also the first person to achieve the then world-record lowest temperature of 1.5 K.

Onnes began investigating the resistance of pure metals at very low temperatures. Some scientists believed at the time that the resistance of metals would greatly increase, or even become infinite near absolute zero (–273°C). Other scientists, Onnes among them, believed that the electrical resistance would eventually drop to nil.

In 1911, Onnes immersed a solid wire of mercury into liquid helium and found that at 4.2 K its resistance was indeed zero. He called this the *superconducting state*, a new state of matter. Theoretically, once an electric current was started in a loop maintained at superconducting temperatures, it would circulate indefinitely. Onnes was awarded the Nobel Prize in Physics in 1913.

Other metals were soon found to become superconductors at extremely low temperatures—for example, aluminium at 1.2 K and lead at 7.9 K. However, not all metals can become superconducting. Perhaps surprisingly, the excellent electrical conductor copper does not become superconducting at any temperature.

The mechanism by which a metal's electrical resistance drops to zero was not understood at the time. In 1957 three American physicists, John Bardeen, Leon Cooper and Robert Schrieffer, published a paper that explained the phenomenon. The explanation is now known as the BCS theory and required quantum mechanics, not known in Onnes' era. Bardeen, Cooper and Schrieffer were awarded the 1972 Nobel Prize in Physics.

Superconductivity hit the headlines again in the late 1980s when some ceramic compounds were discovered to become superconducting at relatively high temperatures. Despite the fact that copper by itself is not a superconductor, most high-temperature superconductors (HTS) are compounds of copper. An example is the compound yttrium barium copper oxide, $YBa_2Cu_3O_7$, which becomes superconducting at liquid nitrogen temperatures (77 K or –196°C). This new class of ceramic superconductors generated a great deal of interest, and currently the record stands at 135 K (–138°C) using the compound mercury barium calcium copper oxide, $HgBa_2Ca_2Cu_3O_8$. The mechanism by which this class of ceramic materials becomes superconducting is not well understood and so they are currently a very active area of both theoretical and experimental research.

The eventual goal is to discover a material that will become superconducting at room temperature. Such a material would have many applications and would not require all the equipment and expense to maintain extremely low or high temperatures.

Superconductors have made it possible to produce extremely powerful magnets. Such magnets find applications in magnetic resonance imaging (MRI) machines used in medical scanning, in maglev (magnetic levitation) trains that float above the rails and therefore do not experience friction and clunking from the rails, and in the Large Hadron Collider to bend extremely high speed protons around the beam-line.

FIGURE 4.1.13 The Dutch physicist Kamerlingh Onnes.

RESISTORS AND POWER

A particular combination of resistors will draw different amounts of power depending on whether the resistors are wired in series or parallel. In general, since resistors in parallel circuits will draw more current than resistors in series circuits, parallel circuits use more power than series circuits containing the same resistors.

Recall from Chapter 3 that the equation for power is:

$P = V \times I$

where P is the power (W)

V is the voltage (V)

I is the current (A).

Worked example 4.1.6

COMPARING POWER IN SERIES AND PARALLEL CIRCUITS

Consider a 100 Ω and a 300 Ω resistor wired in parallel with a 12 V cell. Calculate the power drawn by these resistors. Compare this to the power drawn by the same two resistors when wired in series.	
Thinking	**Working**
Calculate the equivalent resistance for the parallel circuit.	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$ $= \frac{1}{100} + \frac{1}{300}$ $= \frac{3}{300} + \frac{1}{300}$ $= \frac{4}{300}$ $R_T = \frac{300}{4}$ $= 75\ \Omega$
Calculate the total current drawn by the parallel circuit.	$V = IR$ $\therefore I = \frac{V}{R} = \frac{12}{75} = 0.16$ A
Use the power equation to calculate the power drawn by the parallel circuit.	$P = VI$ $= 12 \times 0.16 = 1.92$ W
Calculate the equivalent resistance for the series circuit.	$R_T = R_1 + R_2 + \ldots + R_n$ $= 100 + 300$ $= 400\ \Omega$
Calculate the total current drawn by the series circuit.	$V = IR$ $\therefore I = \frac{V}{R} = \frac{12}{400} = 0.03$ A
Use the power equation to calculate the power drawn by the series circuit.	$P = VI$ $= 12 \times 0.03 = 0.36$ W
Compare the power drawn by the two circuits.	$\frac{P_{\text{parallel}}}{P_{\text{series}}} = \frac{1.92}{0.36} = 5.33$ The parallel circuit draws over 5 times as much power as the series circuit.

Worked example: Try yourself 4.1.6

COMPARING POWER IN SERIES AND PARALLEL CIRCUITS

Consider a 200 Ω and a 800 Ω resistor wired in parallel with a 12 V cell. Calculate the power drawn by these resistors. Compare this to the power drawn by the same two resistors when wired in series.

PHYSICS IN ACTION

High power–low power

Simple heaters of various sorts often have a 'three heat' switch. An electric blanket will usually have 'low', 'medium' and 'high' settings, for example. Rather than making three different heating elements, the manufacturer can use two elements in different series and parallel combinations to obtain the three heat settings. If the two elements are placed in series the total resistance is relatively high and therefore the power will be a minimum, as $P = \frac{V^2}{R}$. For the medium setting one of the elements will be used by itself. The high setting is then achieved by placing both elements in parallel.

It is a simple matter to work out the relative power being used for the three settings. If it is assumed that the resistance of both elements is the same (R) and does not change appreciably with temperature, the effective resistance in the three cases will be given by:

Low heat (two elements in series): $R_T = R + R = 2R$

Medium heat (one element only): $R_T = R$

High heat (both in parallel): $R_T = \frac{1}{2}R$.

As the power is inversely proportional to the resistance $\left(P = \frac{V^2}{R}\right)$, if we call the high setting 100%, then the others will be 50% and 25%.

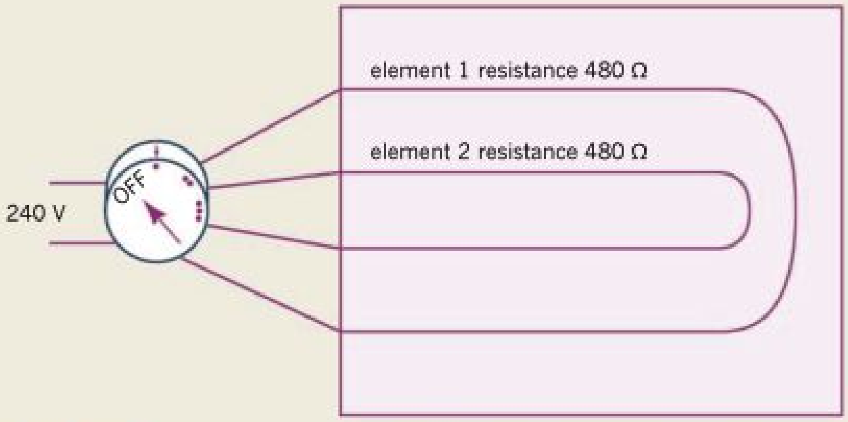


FIGURE 4.1.14 An example of the use of series and parallel combinations of resistors to achieve three heat settings for an electric blanket.
To obtain the various settings the following circuits are used:
OFF Not connected to power
• Resistors connected in series, $R = 960\ \Omega$, $I = 0.25$ A, $P = 60$ W
• • Only one resistive element is connected, $R = 480\ \Omega$, $I = 0.5$ A, $P = 120$ W
• • • Resistors connected in parallel, $R = 240\ \Omega$, $I = 1.0$ A, $P = 240$ W

PHYSICS IN ACTION

Parallel connections

All household appliances and lights are connected in parallel. This is done for two reasons.

Figure 4.1.15 shows a TV, air-conditioner, heater and washing machine connected in series. Each of these devices is designed to operate at 240 V.

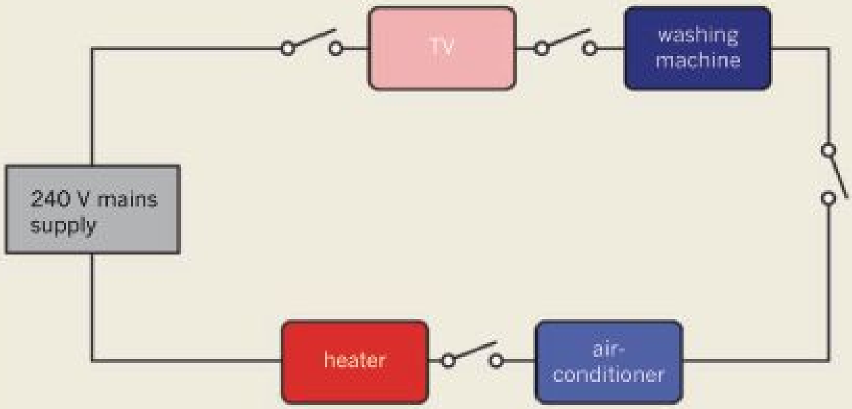


FIGURE 4.1.15 Household appliances connected in series.

A circuit designed like the one in Figure 4.1.15 poses many problems. Firstly, all of the devices need to be switched on for the circuit to operate.

Secondly, the 240 V supplied to the circuit needs to be shared among all the components. Each component in the circuit would receive far less than the 240 V they require to operate. Also, as more and more devices are added to the circuit, the share of the 240 V would become smaller. This system could never be practical.

The circuit diagram in Figure 4.1.16 shows how the same devices could be connected in parallel.

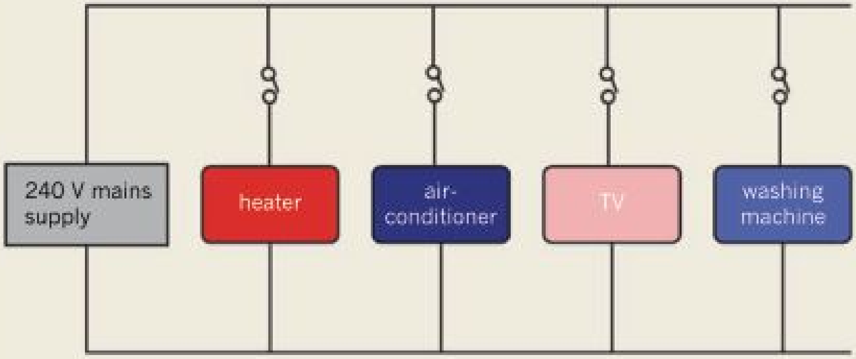


FIGURE 4.1.16 Household appliances connected in parallel.

Each device in the parallel circuit receives the same voltage, 240 V. Each device can be independently switched on or off without affecting the others and more devices can be added to this system without affecting the operation of the others.

4.1 Review

SUMMARY

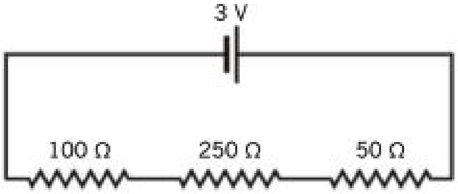
- When resistors are connected in series, the:
  - current through each resistor is the same
  - sum of the potential differences is equal to the potential difference provided to the circuit
  - equivalent effective resistance R_T is equal to the sum of the individual resistances.
- Parallel circuits allow individual components to be switched on and off independently.
- When resistors are connected in parallel, the:
  - voltage across each resistor is the same
  - current is shared between the resistors
  - equivalent effective resistance is given by the equation:
    $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$$
- Complex circuit analysis may require the calculation of both equivalent series and equivalent parallel resistances.
- A parallel circuit generally draws more power than a series circuit using the same resistors.

KEY QUESTIONS

1 Two 20 Ω resistors are connected in series with a 6 V battery. What is the voltage drop across each resistor?

A 0.3 V
B 3 V
C 6 V
D 12 V

2 Consider the series circuit below.



a Calculate the current flowing in the circuit. Give your answer correct to two significant figures.
b Calculate the potential difference across the 100 Ω resistor in the circuit.

3 Two equal resistors are connected in parallel and are found to have an equivalent resistance of 68 Ω. Calculate the resistance of each resistor.

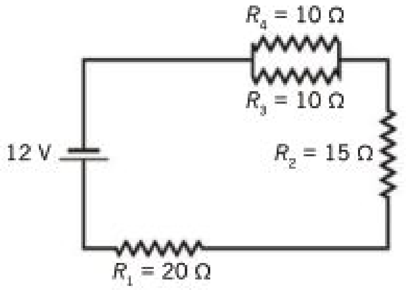
4 A 20 Ω resistor and a 10 Ω resistor are connected in parallel to a 5 V battery. Give your answers correct to two decimal places.

a Calculate the current drawn from the battery.
b Calculate the current flowing through the 20 Ω resistor.
c Calculate the current flowing through the 10 Ω resistor.

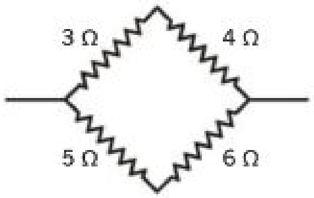
5 A 40 Ω resistor and a 60 Ω resistor are connected in parallel to a battery, with 300 mA flowing through the 40 Ω resistor.

a Calculate the voltage of the battery.
b Calculate the current flowing through the 60 Ω resistor.

6 Calculate the potential difference across, and the current through, each resistor in the circuit below.



7 Calculate the equivalent resistance of the combination of resistors shown below.



8 Four 20 Ω light bulbs are connected to a 10 V battery. What is the total power output of the circuit if the light bulbs are connected:

a in series?
b in parallel?

9 Why are household circuits wired in parallel?

A to reduce the amount of expensive copper wire used
B to reduce the amount of current drawn by the household
C to allow appliances to be switched on and off independently
D to reduce the amount of electrical energy used by the household

4.2 Using electricity

Electricity is a convenient and versatile form of energy. The electrical appliances in homes transform electrical energy into other forms. Devices such as a television, laptop, tablet computer or mobile phone contain complex electronic circuitry. However, the components that make up those circuits are reasonably simple. Understanding how the individual components work is essential to understanding how more complex circuits operate.

TRANSDUCERS

A **transducer** is a device that receives a signal in the form of one type of energy and converts it into another form of energy. For example, a microphone is a transducer that converts sound energy into an electrical signal. A microphone is a type of input transducer because it takes one form of energy (sound) and converts it into electrical energy. An output transducer is one that converts the electrical energy back into another form of energy.

Most complex electrical circuits can be understood as a combination of input and output transducers separated by some form of signal-processing circuitry. The flowchart in Figure 4.2.1 explains the process in a simple way.

FIGURE 4.2.1 An electrical circuit can be modelled as an input transducer followed by signal processing circuitry and then an output transducer.

For example, the input transducer could be a simple battery with a switch while the output transducer is a light globe. The wires connecting the battery, switch and light bulb are the signal-processing circuitry; in this case, the signal is simply carried from one transducer to another without really being processed.

Other, more complex, circuits may use sophisticated input transducers that can produce a range of electrical outputs. The signal-processing circuitry may amplify the signal or change it in some way while the output transducer could produce light, sound or almost any other type of energy imaginable.

Signal-processing components

Potentiometer

FIGURE 4.2.2 Different types of potentiometer.

One of the simplest forms of signal-processing is to vary the amount of current or voltage in a circuit using a variable resistor or **potentiometer**. Although potentiometers can take a variety of forms (as seen in Figure 4.2.2), the basic design is always the same, consisting of a three-terminal resistor with a sliding or rotating contact called the wiper.

If the potentiometer is connected using just one end and the wiper, it acts as a simple variable resistor: the further the wiper slides or rotates, the greater the resistance value.

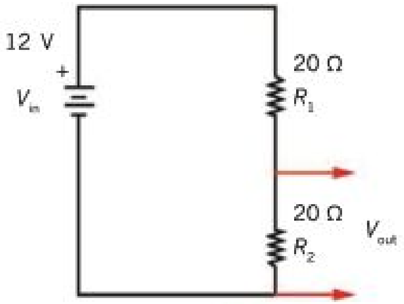
A potentiometer can also be used to divide voltage. This means that if a potential difference is applied across the two ends, then the wiper can be used to access any voltage between these two extremes. For example, if the ends of a potentiometer were connected to a 12 V battery, then setting the wiper in the middle position would give a 6 V difference between it and either end, i.e. it would divide the total resistance of the potentiometer in half.

Just as a potentiometer can divide voltage, so too can a **voltage divider** circuit. A voltage divider circuit, like the one shown in Worked example 4.2.1 on page 139, is simply a series circuit with two or more components. The circuit is called a voltage divider because the voltage supplied to the circuit is shared (or divided) between the components in the circuit. While in this worked example the components are two fixed resistors, these could be replaced with any of the variable resistors described in this section.

Worked example 4.2.1

VOLTAGE DIVIDER

A voltage divider is constructed from a 12 V battery and two 20 Ω resistors as shown. Calculate the voltage output, V_{out}, of the circuit.

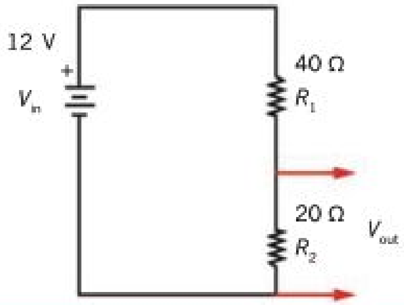


Thinking	Working
Calculate the total resistance of the circuit.	$R_T = 20 + 20$ $= 40\ \Omega$
Calculate the current flowing through the circuit.	$I = \frac{V}{R}$ $= \frac{12}{40}$ $= 0.3$ A
Calculate the potential difference across the second resistor.	$V_{out} = IR$ $= 0.3 \times 20$ $= 6$ V

Worked example: Try yourself 4.2.1

VOLTAGE DIVIDER

A voltage divider is constructed from a 12 V battery, a 40 Ω resistor and a 20 Ω resistor as shown. Calculate the voltage output, V_{out}, of the circuit.



Potentiometers are widely used as light dimmer switches and for volume control on sound systems.

Consider Worked example 4.2.1 above. Each of the resistors provides exactly the same amount of resistance to the circuit. They divide the supply voltage exactly in half. If instead, one of the resistors was larger than the other, the potential difference (voltage drop) would be greater over the larger resistor.

If you completed Worked example: Try yourself 4.2.1, you would have seen that the voltage was not divided equally between the resistors as they were not equal in size.

If one of the fixed resistors is replaced with a variable resistor, it is possible to constantly change the way the supply voltage is divided. This is what happens in a dimmer switch.

Figure 4.2.3 shows what happens to the brightness of a light globe connected in series with a potentiometer set at low resistance and then at high resistance. When the potentiometer is set to its lowest resistance, the potential difference across the globe will be large and the bulb will glow brightly. When the resistance of the potentiometer increases, the potential difference across the light globe decreases and the bulb is dim.

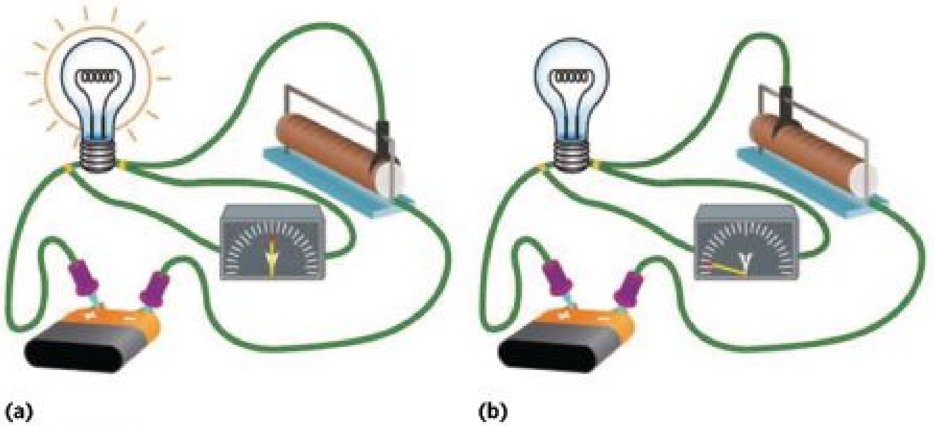


FIGURE 4.2.3 (a) When the potentiometer is at its lowest resistance, the globe glows brightly. (b) As the resistance of the potentiometer increases, the globe glows less brightly.

It is important to note that, along with voltage, the current in a circuit is also affected by changing the resistance of the potentiometer. In the circuit shown in Figure 4.2.3(b), when the resistance of the potentiometer is very high, the current flowing in the circuit decreases.

EXTENSION

Voltage divider formula

Circuit analysis shows that the output voltage, V_{out}, of a voltage divider can be given by a generalised formula. Consider the circuit in Figure 4.2.4.

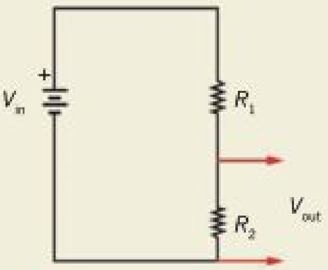


FIGURE 4.2.4 A circuit diagram for a generalised voltage divider.

Since the equivalent series resistance of the circuit is given by $R_T = R_1 + R_2$, the current in the circuit is given by:

$$I = \frac{V}{R} = \frac{V_{in}}{R_1 + R_2}$$

Using Ohm's law,

$$V_{out} = IR = \frac{V_{in}}{R_1 + R_2} \times R_2$$

Alternatively,

$$V_{out} = V_{in} \times \frac{R_2}{R_1 + R_2}$$

Input transducers

Thermistor

A **thermistor**, such as the one shown in Figure 4.2.5, is a type of variable resistor. The resistance of the thermistor, however, depends on its temperature.

FIGURE 4.2.5 The resistance of a thermistor changes with its temperature.

Thermistors are usually categorised as either NTC (negative temperature coefficient) or PTC (positive temperature coefficient).

As the labels imply, an NTC thermistor's resistance decreases as its temperature increases. A PTC thermistor's resistance increases as its temperature increases. An incandescent light globe is an example of a PTC thermistor. As its filament gets hotter with increasing current, its resistance increases.

Figures 4.2.6(a) and Figure 4.2.6(b) show how the resistances of these devices varies with temperature. In comparison, Figure 4.2.6(c) shows how a fixed (non-variable) resistor, in ideal conditions, will maintain constant resistance even when the temperature changes.

NTC thermistors are semiconductor devices and are the most common type used in applications where electronic control is needed in response to temperature change.

Thermistors are used in a wide variety of temperature-control applications such as in refrigerators, toasters, coffee-makers, electrical circuits or engines.

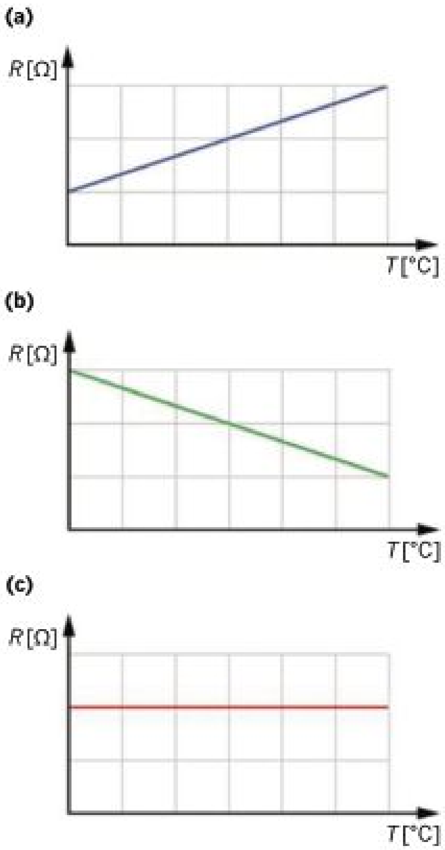


FIGURE 4.2.6 (a) The increase in resistance of a light globe with increasing temperature. (b) The decrease in resistance of a NTC thermistor with increasing temperature. (c) The independence of an ideal resistor's resistance with temperature.

Transducers in voltage divider circuits

Voltage divider circuits, like the one in Worked example 4.2.1, can contain fixed or variable resistors. Page 139–140 described how a potentiometer used in this kind of circuit could act as a dimmer switch.

Thermistors can also be connected in voltage divider circuits with other resistors. To know how a thermistor changes its resistance with temperature, manufacturers provide a characteristic curve. Figure 4.2.7 shows the characteristic curve for a particular thermistor and also how the thermistor can be connected in a voltage divider circuit.

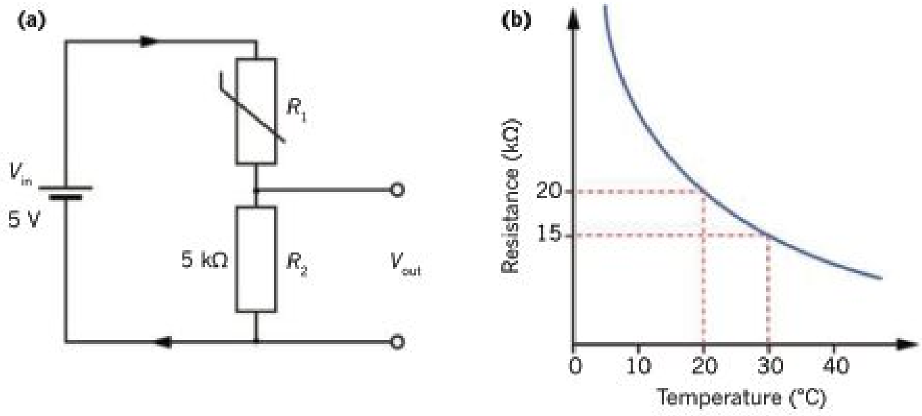


FIGURE 4.2.7 (a) A thermistor (R_1) is connected in series with a fixed resistor (R_2). This is a voltage divider circuit as the voltage supplied by the battery will be divided between the two resistors. (b) The characteristic curve of the thermistor shows how its resistance varies with temperature.

Worked example 4.2.2

THERMISTOR IN A VOLTAGE DIVIDER CIRCUIT

A voltage divider circuit includes a thermistor (R_1) and a fixed resistor (R_2). The characteristic curve of the thermistor and the circuit are shown in Figure 4.2.7 on page 141. Using the graph and the information included on the circuit diagram, determine the following:

a the resistance of the thermistor at 30°C

Thinking	Working
The resistance of the thermistor can be read straight from the graph at the point where the temperature is 30°C.	$R = 15\ \text{k}\Omega$

b the current in the circuit

Thinking	Working
Find the effective resistance of the circuit. Note that the fixed resistor is 5 kΩ.	$R_T = 15\ \text{k}\Omega + 5\ \text{k}\Omega$ $= 20\ \text{k}\Omega$
Find the current.	$I = \frac{V}{R}$ $= \frac{5}{20000}$ $= 0.25\ \text{mA}$

c the output potential difference, V_{out}.

Thinking	Working
Use Ohm's law to calculate the voltage across the fixed resistor.	$V = IR$ $= 0.00025 \times 5000$ $= 1.25\ \text{V}$

Worked example: Try yourself 4.2.2

THERMISTOR IN A VOLTAGE DIVIDER CIRCUIT

A voltage divider circuit includes a thermistor (R_1) and a fixed resistor (R_2). The characteristic curve of the thermistor and the circuit are shown in Figure 4.2.7. Using the graph and the information included on the circuit diagram, determine the following:

a the resistance of the thermistor at 20°C

b the current in the circuit

c the output potential difference, V_{out}.

The circuit shown in Figure 4.2.7(a) can be used as a temperature sensor. When the voltage over the fixed resistor reaches a given value, it can signal for a heater or cooling system to switch on or off.

Light-dependent resistor

The device shown in Figure 4.2.8 is an input transducer that provides a variable resistance in the circuit. As its name suggests, a **light-dependent resistor** (also known as an LDR or photoresistor) is a transducer whose resistance depends on the amount of light falling on it. Usually an LDR is designed to have a very high resistance in the dark (for example a few million ohms). In the light, their resistance can drop as low as a few hundred ohms.

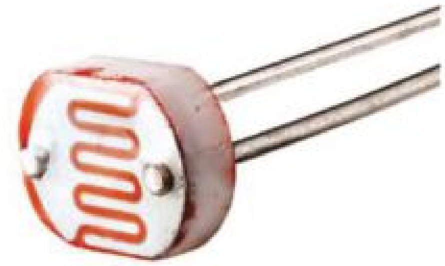


FIGURE 4.2.8 The resistance of a light-dependent resistor changes depending on how much light is falling on it.

As the resistance of the LDR changes it has the effect of changing the potential difference (voltage drop) as well. As the LDR's resistance increases, so does the potential difference across it. A decrease in resistance means the voltage drop decreases as well. This is because in any device with high resistance, more voltage (electrical potential energy) is needed to make the charges flow through the device.

Increasing the resistance of a device also has an effect on the current in the circuit. It is much harder for charges to flow through a device with a high resistance. So, as the resistance of the LDR increases, the current flowing through it decreases. As current is the same at every point of a series circuit, the current in the entire circuit will decrease.

LDRs have a range of applications in devices such as camera light meters, street lights and night lights. They can also be used to detect when a person's body blocks a beam of light shining across a doorway, triggering a buzzer or doorbell chime.

The circuit diagrams in Figure 4.2.9 show how an LDR can be connected in a circuit to achieve different outputs. When the LDR and the light globes are connected in series, the bulb switches on only if light falls on the LDR. This is because, as they are connected in series, current must flow through the LDR in order for it to also flow through the bulb. The only way current will flow through the LDR is if its resistance is low, and this happens when light falls upon it.

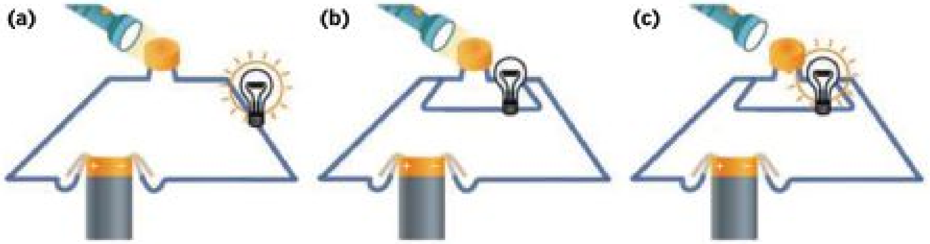


FIGURE 4.2.9 (a) Connected in series, a light globe lights up when the LDR is illuminated. (b) and (c) Connected in parallel, a light globe lights up when the LDR is not illuminated.

When the LDR and the light globe are connected in parallel, the bulb switches on only if the LDR is not illuminated (lit up). The current divides between the branches of parallel circuits but if one path has very high resistance, the charges will 'choose' the other path, of least resistance. When the LDR is not illuminated, it has very high resistance and so the current passes through the other branch, lighting the bulb.

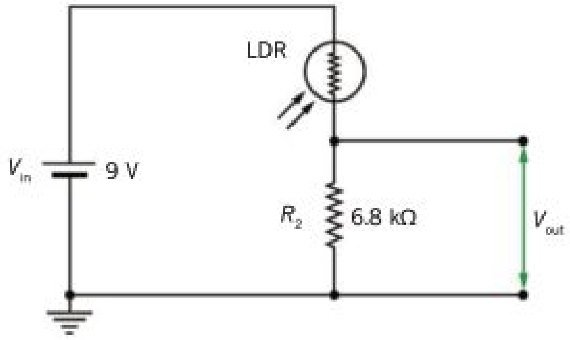
It is more likely that an LDR is used in a circuit where a light globe switches on when the ambient (surrounding) light is low. In these cases, the LDR needs to be connected in parallel with the bulb.

When provided with a characteristic curve, the operation of an LDR in a voltage divider circuit can be analysed. The characteristic curve provides information about the resistance of the LDR, and then using Ohm's law, current and voltage can be calculated.

Worked example 4.2.3

LDR IN A VOLTAGE DIVIDER CIRCUIT

A voltage divider circuit includes an LDR (R_1) and a fixed resistor (R_2).



Using the information included on the circuit diagram and the fact that the LDR has a resistance of 20 kΩ at a light intensity of 100 mW m^{-2}, determine the following:

a the total resistance of the circuit

Thinking	Working
The total resistance equals the resistance of the LDR plus the 6.8 kΩ fixed resistor.	$R_T = 20 + 6.8 = 26.8$ kΩ

b the current in the circuit

Thinking	Working
Find the current using Ohm's law.	$I = \frac{V}{R}$ $= \frac{9}{26\,800}$ $= 0.34$ mA

c the output potential difference, V_{out}.

Thinking	Working
Use Ohm's law to calculate the voltage across the fixed resistor.	$V = IR$ $= 0.00034 \times 6800$ $= 2.3$ V

Worked example: Try yourself 4.2.3

LDR IN A VOLTAGE DIVIDER CIRCUIT

A voltage divider circuit includes an LDR (R_1) and a fixed resistor (R_2).

Using the information included on the circuit diagram in Worked example 4.2.3 and the fact that the LDR has a resistance of 3.0 kΩ at a light intensity of 2000 mW m^{-2}, determine the following:

a the total resistance of the circuit

b the current in the circuit

c the output potential difference, V_{out}.

Diodes

A **diode** is a semiconductor device that has the special property that it will only allow electrical current to flow through it in one direction. Several types of diodes are shown in Figure 4.2.10. This simple function has a wide range of applications. Diodes are often used to protect current-sensitive components or circuits from stray currents. They are also used in circuits which convert **alternating current** (AC) into **direct current** (DC). This is a particularly important application since the household mains supply (available through a power point) is alternating current while most electronic devices operate on direct current.

It is important to note that diodes are non-ohmic, that is, their *I–V* graphs are not linear (a straight line). The resistance across a diode is not constant for all potential differences. Figure 3.4.6(b) on page 115 shows a typical *I–V* graph for a diode.

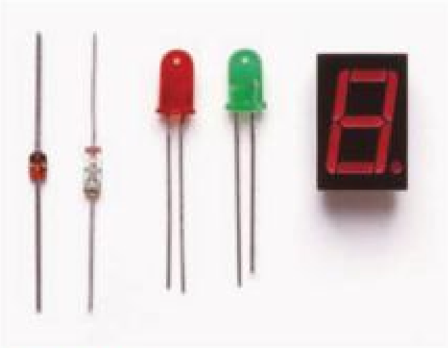
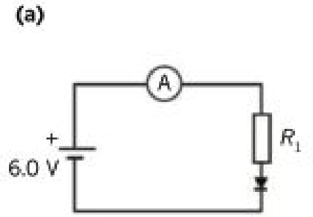


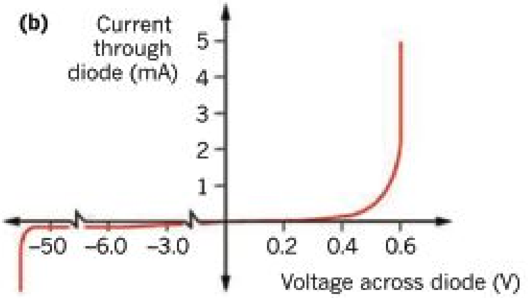
FIGURE 4.2.10 Different types of diodes.

Worked example 4.2.4

DIODES

A student is investigating the current–voltage characteristics of a diode using the circuit shown in diagram (a). The *I–V* graph for this diode is illustrated in (b). The current in the circuit is measured as 4.5 mA.





a What is the potential difference across the diode?

Thinking	Working
The potential difference across the diode can be read directly from the *I–V* graph.	At 4.5 mA the voltage across the diode will be 0.6 V.

b What is the value of R_1?

Thinking	Working
The battery voltage is 6.0 V, the voltage drop across the diode is 0.6 V and the current is 4.5 mA. Use this information to find R_1.	Voltage across $R_1 = 6.0 - 0.6$ $= 5.4$ V Resistance $R_1 = \frac{V}{I}$ $= \frac{5.4}{4.5 \times 10^{-3}}$ $= 1200\ \Omega$

Worked example: Try yourself 4.2.4

DIODES

A student is investigating the current–voltage characteristics of a diode using the same circuit diagram and *I–V* graph shown in Worked example 4.2.4.

a What is the potential difference across the diode?

b What is the value of R_1?

Output transducers

Light-emitting diode

A **light-emitting diode** or LED is a type of diode that emits light as current passes through it. LEDs are very popular in electronic circuits because they can operate using much smaller currents and voltages than incandescent light globes. Other advantages of LEDs over light globes are their small size, their very long life span and their low price.

Although the most common colour for an LED is red, they can be designed to produce light at almost any wavelength in the visible or near-visible spectrum. Some available colours for LEDs are shown in Figure 4.2.11.

FIGURE 4.2.11 Different coloured LEDs.

Recently, LEDs have been increasingly used outside the field of electronics as highly energy-efficient replacements for traditional incandescent light globes. For example, the LED downlights shown in Figure 4.2.12 are quickly replacing incandescent and halogen lighting traditionally used in homes.

FIGURE 4.2.12 A group of LEDs can be used to produce the same amount of light as a traditional incandescent light globe, but use much less energy to produce it.

Many of the common LEDs have a switch-on voltage of around 1.8–3.5 V. The switch-on voltage of a diode is the voltage at which the current flowing through the diode increases very rapidly. Figure 4.2.13 shows a typical circuit used to activate an LED. Note that the limiting resistor is used to keep the LED's forward-biased current within the maximum allowed rating. If the current exceeds this limit, the LED will be permanently damaged.

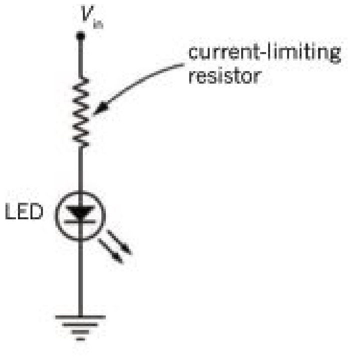


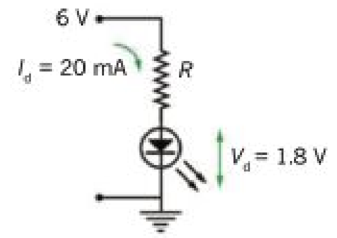
FIGURE 4.2.13 An LED driver circuit.

Worked example 4.2.5

LEDs

An LED has the following optimum operating characteristics:

$I_d = 20$ mA when $V_d = 1.8$ V



Determine the value of the current-limiting resistor (R) if the current through the diode is to be limited to 20 mA when powered by a 6 V battery

Thinking	**Working**
The 6 V is divided between the resistor R and the LED. Work out the voltage drop across R.	$V_R = 6 - 1.8 = 4.2$ V
Use Ohm's law to calculate the value of R.	$R = \frac{V}{I}$ $= \frac{4.2}{20 \times 10^{-3}}$ $= 210\ \Omega$

Worked example: Try yourself 4.2.5

LEDs

An LED has the following optimum operating characteristics:

$I_d = 30$ mA when $V_d = 1.9$ V.

Use the diagram from Worked example 4.2.5 to determine the value of the current-limiting resistor (R) if the current through the diode is to be limited to 30 mA when powered by a 6 V battery.

PHYSICSFILE

Diodes and LEDs

Diodes and LEDs only allow current to pass in one direction. As such it is essential they are placed in circuits the correct way around. When a diode is placed in the circuit in the correct way, it is said to be forward biased. If a diode is connected the wrong way around it will shut down part or all of a circuit. Therefore manufacturers are careful to make it clear which way these components are to be orientated.

The circuit symbol for a diode consists of a triangle and a line. This can be seen in Figure 4.2.14. Figure 4.1.15 shows the circuit symbol of an LED. The direction of the arrows shows the only direction that conventional current will flow through the devices.

FIGURE 4.2.14 The red arrow under the diode symbol shows the direction of current flow.

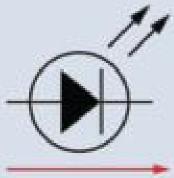


FIGURE 4.2.15 An LED circuit symbol. Current will flow the through the LED in the direction given by the red arrow.

4.2 Review

SUMMARY

- Most complex electrical circuits can be understood as a combination of input and output transducers separated by signal-processing circuitry.
- The basic design of potentiometers is always the same, consisting of a three-terminal resistor with a sliding or rotating contact called the wiper.
- A potentiometer can also be used to divide voltage. Potentiometers are widely used as light dimmer switches and for volume control on sound systems.
- A voltage divider circuit is simply a series circuit with two or more components. The circuit is called a voltage divider because the voltage supplied to the circuit is shared (or divided) between the components in the circuit.
- The resistance of the thermistor depends on its temperature. Thermistors are used in a wide variety of temperature-control applications such as in refrigerators, toasters, coffee makers, electrical circuits or engines.
- A light-dependent resistor is a transducer whose resistance depends on the amount of light falling on it. Usually an LDR is designed to have a very high resistance in the dark and a very low resistance in the light.
- LDRs have a range of applications. They can be found in camera light meters, street lights and night lights. They can also be used to detect when a person's body blocks a beam of light shining across a doorway, triggering a buzzer or doorbell chime.
- A diode is a semiconductor device that will only allow electrical current to flow in one direction through it. They are used in circuits which convert alternating current into direct current.
- A light-emitting diode or LED is a type of diode that emits light as current passes through it.
- LEDs are increasingly used as highly energy-efficient replacements for traditional incandescent light globes.

KEY QUESTIONS

1 A 10 V battery is connected to a voltage divider circuit which consists of a 100 Ω resistor and a 300 Ω resistor.

a What is the voltage across the 300 Ω resistor?

A 2.5 V
B 3.3 V
C 7.5 V
D 13.3 V

b What is the voltage drop across the 100 Ω resistor?

A 2.5 V
B 3.3 V
C 7.5 V
D 13.3 V

2 A 24 V battery is connected to a voltage divider circuit consisting of a 1 kΩ resistor and a 2 kΩ resistor. What is the voltage across the 2 kΩ resistor?

3 Copy the table below and classify the following components under the appropriate heading:
diode, LED, LDR, light globe, microphone, potentiometer, speaker, thermistor

Input transducer	Signal-processing component	Output transducer

4 Which of the following components only allow current to flow in one direction? (More than one correct answer is possible.)

A diode
B LED
C potentiometer
D thermistor

5 As the temperature of a PTC thermistor increases, what happens to its resistance?

A decreases
B increases
C remains constant
D approaches infinity

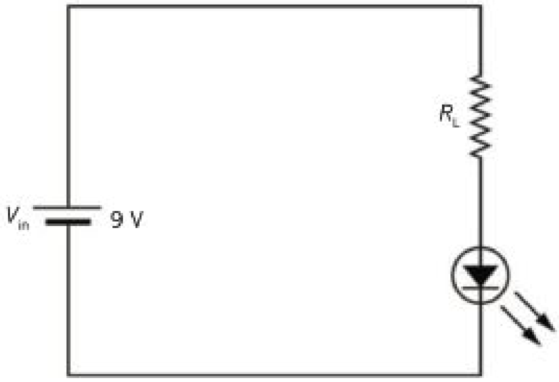
6 Which of the following could be used as an input transducer in a circuit to control the temperature inside a refrigerator?

A LED
B thermistor
C LDR
D diode

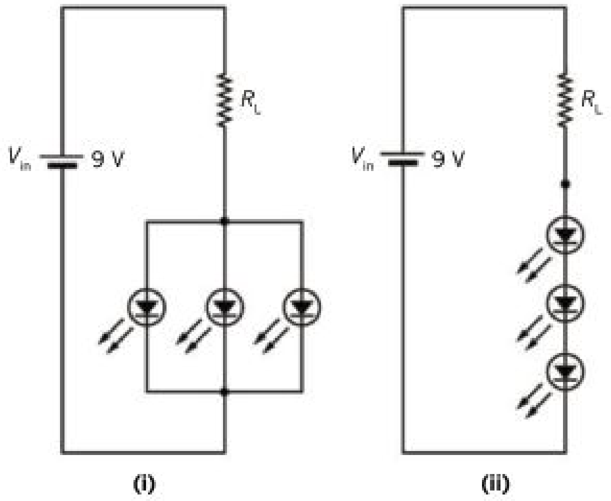
7 As the amount of light falling on an LDR increases, what happens to the resistance of the LDR?

A decreases
B increases
C remains constant
D approaches infinity

8 The LED in the circuit below has a switch-on voltage (V_s) of 2.0 V and an operating current of 20 mA. Determine the value of R_L for the LED to be operating correctly.



9 All LEDs in the circuits (i) and (ii) below are identical. Each LED has a switch-on voltage (V_s) of 2 V and draws a current of 20 mA for optimal light production. (If the current is much smaller, the LED light output is too dim. If the current is much larger, the LED overheats and burns out.)



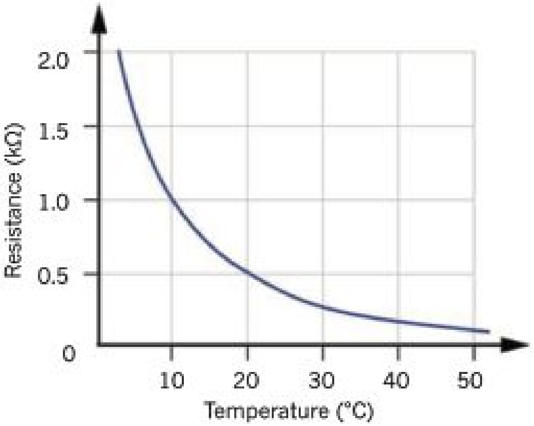
(i) **(ii)**

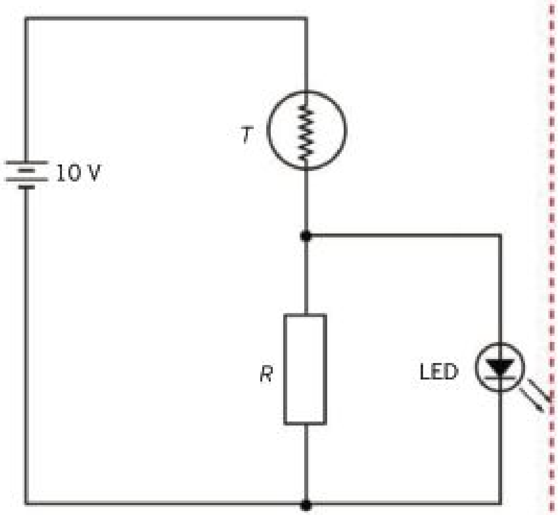
a For each circuit determine the R_L that gives optimum operation for all of the LEDs.

b Which combination of LEDs (circuit (i) or (ii)) emits the most light with these particular values of R_L?

c Assuming circuits (i) and (ii) have exactly the same ideal battery power supply, determine which circuit will emit light (i.e. with the LEDs operating at their optimum level) for the longest time. Explain your answer.

10 The following graph shows the resistance–temperature characteristics of a thermistor. A circuit uses this thermistor as part of a temperature sensor which can activate an LED whenever the temperature rises above a certain level.





a What is the resistance of the thermistor at 20°C?

b The potential difference across the LED at 20°C is 2.5 V and the current through it is 11 mA. What is the value of R?

c The LED is activated by a minimum potential difference of 2.0 V across it which gives a current through it of 4.8 mA. What is the minimum temperature that will activate the LED?

4.3 Electrical safety

Homes, schools and workplaces are filled with all sorts of electrical appliances. You use these appliances every day. A scientific understanding of electricity can help you to understand how they work and how to make sure you use them safely and effectively.

ELECTRICITY IN THE HOME

Circuits in the home

The wiring for a house is much more complicated than the relatively simple series and parallel circuits considered so far. Figure 4.3.1 shows the basic structure of the electrical wiring in a house. Most appliances and power points are wired in parallel to allow them to be switched on and off independently of each other.

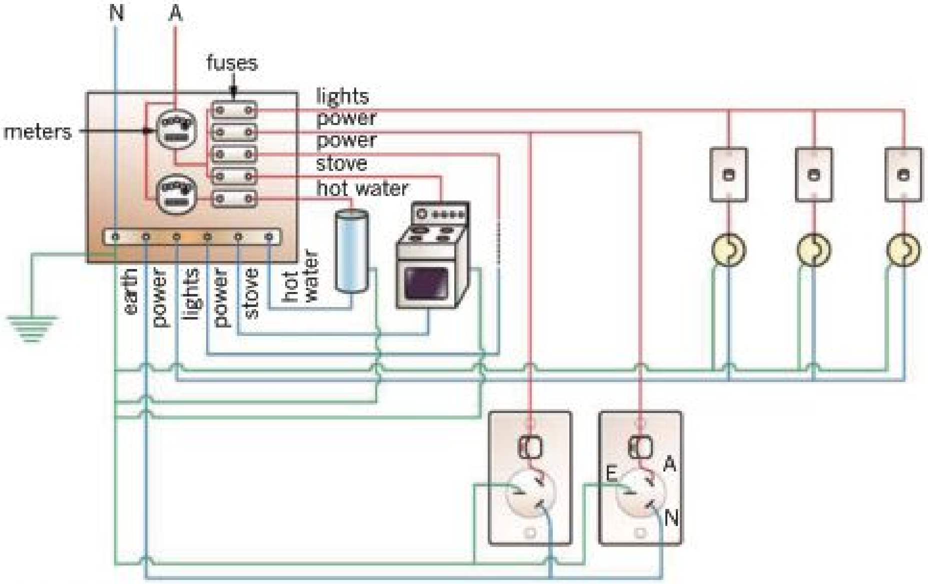


FIGURE 4.3.1 A household wiring diagram includes active (A), neutral (N) and earth (E) wires.

Power points in Australia have three pins. Each of these pins is connected to a different wire. Figure 4.3.1 shows the arrangement of the active, neutral and **earth** pins on a typical power point. The wires carrying the electric current to and from the appliance are known as the active wire (usually red or brown) and the neutral wire (usually black or blue). The third wire is an important safety feature called the earth wire (usually green or green and yellow). Figure 4.3.2 shows the corresponding active, neutral and earth pins in a power cord.

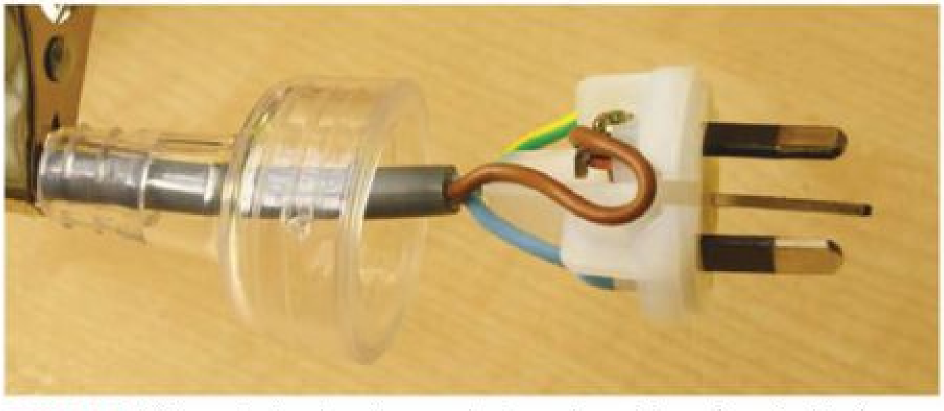


FIGURE 4.3.2 This three-pin plug shows the correct colour code used: brown for active, blue for neutral and green/yellow for earth.

The purpose of household electrical circuits is to enable electrical energy to be transferred to electrical appliances, where it is transformed into a range of other useful forms of energy. For example, an electric oven converts electrical energy into heat, whereas fans convert electrical energy into kinetic energy. Power points give users the option of connecting their own appliances and therefore choosing the type of energy produced.

Figure 4.3.3 shows that, before being distributed to various parts of the house, the active wires pass through meters that measure the amount of electrical energy supplied to the household.

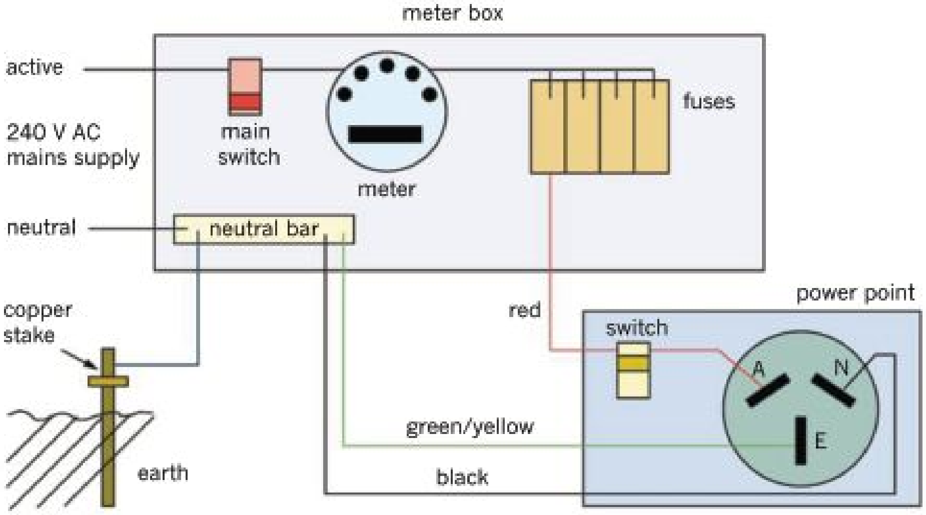


FIGURE 4.3.3 The active wire passes through the meter box before being connected to power points throughout the house. The circuit is completed by the neutral wire, returning through the neutral bar.

AC/DC

The electrical current that comes out of a power point is very different to the current that comes from a battery or electric cell (Figure 4.3.4).

FIGURE 4.3.4 A power point provides alternating current (AC) whereas a car battery provides direct current (DC).

For a start, the potential difference between the active and neutral pins of a household power point is 240 V. This is much higher than most electric cells or batteries can provide. In fact, 240 V is the equivalent of putting twenty 12 V car batteries together in series.

Secondly, because of the way household electrical energy is generated, it is delivered as an alternating current (AC). This means that the electrons in the wire oscillate backwards and forwards in the wire. In comparison, a battery provides direct current (DC), which means the electrons travel directly from one terminal of the battery to the other, in one direction only.

Fortunately, most household AC systems can be modelled using simple DC circuits.

PHYSICSFILE

Kilowatt hours and Joules

The energy unit of kW h can be converted to joules (J) by multiplying the number of kW h by 3 600 000. Therefore 1 kW h = 3.6×10^6 J or 3.6 MJ.

This value is simply calculated as follows:

$$\begin{aligned} 1 \text{ kW h} &= 1000 \text{ W} \times 60 \text{ minutes} \times 60 \text{ seconds} \\ &= 3\,600\,000 \text{ watts} \times \text{seconds} \\ &= 3.6 \times 10^6 \text{ J} \end{aligned}$$

Electricity bills

The energy consumed by a household appliance is measured in joules. It is the power, P, in watts multiplied by the time, t, in seconds.

$E = Pt$

If you have a look at your home electricity bill, you'll see it is measured in **kilowatt hours** (kW h). This gives a convenient number, without scientific notation, that is easy to write on your bill.

To calculate how much an appliance costs to run, multiply its power consumption in kW by the number of hours it runs for. Then take this number in kW h and multiply it by the cost of electricity per kW h. Complete Worked example 4.3.1 below to calculate the cost of running of some household appliances.

Worked example 4.3.1

CALCULATING THE COST OF ELECTRICITY

A 2000 W air conditioner runs for 5 hours. Assume the price for household electricity is 26 cents per kW h. How much would it cost to run this air conditioner for 5 hours?

Thinking	Working
Convert the power consumption of the appliance to kW.	$\frac{2000 \text{ W}}{1000} = 2$ kW
Use the appropriate equation to multiply the power of the appliance in kW by the number of hours it operates.	$E = Pt$ $= 2 \times 5$ $= 10$ kW h
Multiply the number of kW h by the cost per kW h.	Cost = 10×0.26 = \$2.60

Worked example: Try yourself 4.3.1

CALCULATING THE COST OF ELECTRICITY

A 2500 W iron is used for 2.5 hours. Assume the price for household electricity is 26 cents per kW h. How much would it cost (to the nearest cent) to use this iron for 2.5 hours?

ELECTRICAL SAFETY DEVICES

Household electrical wires can carry large amounts of energy. This means that they have the potential to do a lot of harm. The inherent danger associated with the use of electricity can be reduced using various safety devices.

Fuses and circuit breakers

Since wires heat up when current passes through them, there is a limit to how much current the wires in a house or building can safely carry. Household wiring systems are designed to prevent wires from becoming **overloaded**. Appliances that draw a lot of current, such as ovens, hot-water systems and air-conditioners, are put on separate circuits to lights and power points.

Despite these precautions, overloading can still occur, most often due to a **short circuit**. A short circuit occurs when an electric circuit contains very little resistance. This can occur in an electrical appliance when the insulation between the active and neutral wires becomes damaged and these wires are in direct contact. In household circuits, short circuits are always dangerous situations. Large amounts of current means that wires will heat up, causing insulation to melt or catch alight.

An electric current will always take the path of least resistance. The globe in Figure 4.3.5(a) is on because the current has no alternative but to pass through the high resistance of the bulb. The globe in Figure 4.3.5(b) does not work because the closed switch provides a zero-resistance alternative pathway for the current: a short circuit.

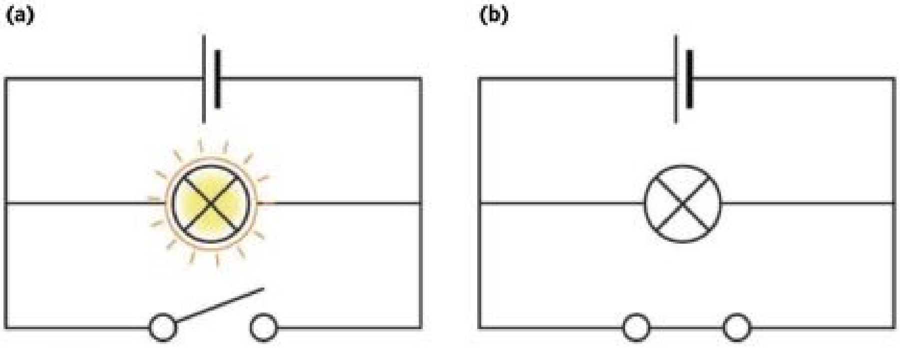


FIGURE 4.3.5 Short circuits. The globe in (a) functions normally, while (b) shows how a short circuit forms through the closed switch, so the bulb does not light.

PHYSICSFILE

Power board overload

Another common reason for circuit overloading is the overuse of power boards and double adaptors. Most power boards are designed to carry a maximum of 10 A of current. If too many high-current appliances, such as heaters, kettles and irons, are plugged into a power board, it can overheat. This may cause the insulation around the wires to melt, causing a short circuit or even a fire.

FIGURE 4.3.6 Overuse of power boards and double adaptors can cause overloading of circuits. This can cause wires and components to overheat and start a fire.

Every domestic electric circuit contains either a **fuse** or a **circuit breaker**. The function of both of these components is to interrupt the flow of current if it exceeds a certain value. Unlike a fuse, a circuit breaker can be easily reset after it has been activated, whereas a fuse needs to be physically replaced once it has melted through. Both fuses and circuit breakers can be chosen for different amounts of current, since some appliances such as ovens and hot-water systems might typically draw much larger amounts of current than regular power points.

Earth wires

Many household electrical appliances such as kettles, toasters and ovens have metal cases. If the active wire inside the appliance becomes loose and touches the case, then the whole case becomes electrically live. If anyone touches the case, the current will flow through their body, with possibly fatal consequences.

To prevent this, an earth wire is permanently connected to the metal case of the appliance, as shown in Figure 4.3.7. When the appliance is plugged in, this wire is connected via the household wiring system to the earth. This means that, if the active wire touches the case, a short circuit will be created and current will immediately flow directly to earth. The large amount of current that flows in this situation should trip the fuse or circuit breaker, alerting users of the appliance to the problem.

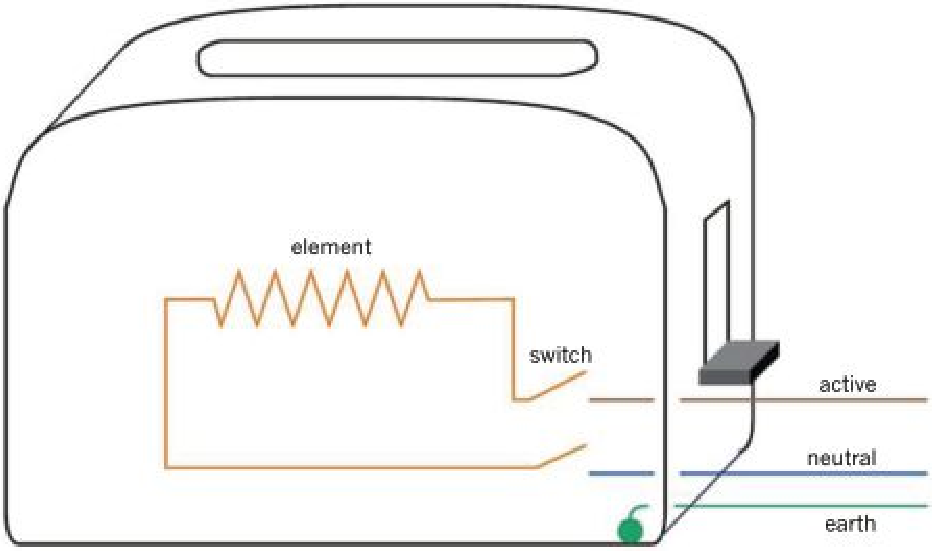


FIGURE 4.3.7 The earth wire inside a metal toaster is permanently connected to the casing.

Double insulation

Double insulation is another way of protecting against the possibility of a loose active wire inside an appliance. This process involves using two insulating barriers to protect users. Often this is done by making the case of the appliance out of plastic. This acts as an insulating layer in addition to the plastic insulation surrounding the active wire inside the appliance. Double-insulated appliances do not need an earth wire, so their electrical plugs only have two pins. They also usually have a special symbol on their cases, indicating that they are double insulated, as shown in Figure 4.3.8.

FIGURE 4.3.8 The double square symbol indicates that this appliance has double insulation and does not use an earth connection.

Residual current devices

Residual current devices, also known as RCDs or earth leakage systems, detect any difference between the current in the active wire and the current in the neutral wire. In a properly operating circuit, these two currents should be exactly the same, but in opposite directions. The most likely reason for a difference between the active and neutral currents is that some current is going to earth through a fault or, in a worse case, through a person. If this happens, the RCD is able to switch off the supply in about 20 milliseconds, hopefully preventing any serious harm.

ELECTRIC SHOCK

Despite all the safety features of modern electrical systems and appliances, each year approximately 50 Australians are killed in electrical accidents. The effect of **electric shock** depends on a number of factors including:

- the amount of current passing through the body
- its duration
- the path it takes through the body.

When attending to a victim of electrocution, it is important to first check if they are still in contact with the electrical source. First aid should only be administered when it is safe to approach the victim (Figure 4.3.9).

FIGURE 4.3.9 The recovery position. Before attempting first aid, make sure that the person is no longer in contact with the source of electric current.

The amount of current that will flow through the body if it is in good contact with a 240 V source is well above the level that will cause death. Anything that improves the contact, like wet hands or bare feet, lowers the resistance and increases the current, potentially to life-threatening levels. Although rubber boots and gloves will increase resistance and lower the current, these should *not* be used as a form of protection in place of other sensible electrical safety precautions.

Since our bodies are relatively poor conductors, electrical energy passing through our bodies is quickly converted into heat and can cause terrible internal and external burns. Table 4.3.1 on page 156 shows the likely effect on the human body of a half-second electric shock at different currents.

Current (mA)	Effect on the body
1	able to be felt
3	easily felt
10	painful
20	muscles paralysed—cannot let go
50	severe shock
90	breathing upset
150	breathing very difficult
200	death likely
500	serious burning, breathing stops, death inevitable

TABLE 4.3.1 The effect of a half-second electric shock on the body.

The amount of electrical energy that enters the body depends on the duration of electrocution. This is why the high voltage spark from a Van de Graaff generator is harmless: the duration of the current is about a microsecond and the total energy delivered is tiny. Table 4.3.2 shows the likely effect on the human body of a 50 mA shock for different time periods.

Time (s)	Effect on the body
less than 0.2	noticeable but usually not dangerous
0.2–4	significant shock, possibly dangerous
more than 4	severe shock, possible death

TABLE 4.3.2 The effect of time on the severity of a shock.

The path that the current takes through the body is also important in determining its effect. Since our bodies are controlled by electrical impulses along the nerves, any current that flows into the body from an external source may interfere with our vital functions. In particular, any current flowing from one arm to the other may cause the chest muscles to contract and breathing to stop. Current through the heart regions can cause the muscles to become uncoordinated and the heart function stops. A brief current of about 80 mA is sufficient to cause fibrillation (irregular contraction of the heart muscle) if it flows directly through the heart. This is the cause of most electrical fatalities.

4.3 Review

SUMMARY

- Electrical energy use in the home is usually measured in kilowatt hours, where 1 kW h = 1000 W × 1 h
- The effect of electrocution depends on the amount of current passing through the body, its duration and the path it takes through the body.
- The danger associated with the use of electricity in the home can be managed by using fuses, circuit breakers, double insulation and residual current devices.

KEY QUESTIONS

1 How does a fuse or circuit breaker increase household electrical safety?

A by breaking the flow of current if it becomes too high

B by breaking the flow of current if a difference is detected between the currents in the active and neutral wires

C by taking current to the earth if the metal casing of the appliance becomes live

D by providing an extra layer of electrical insulation

2 How does double insulation increase household electrical safety?

A by breaking the flow of current if it becomes too high

B by breaking the flow of current if a difference is detected between the currents in the active and neutral wires

C by taking current to the earth if the metal casing of the appliance becomes live

D by providing an extra layer of electrical insulation

3 Convert 10 kW h into J.

4 One of the values given in the information below is incorrect. Use your knowledge of kW h to determine which value it is.

A 750 W air conditioner uses 0.75 kW h of energy in 1 hour. A typical price for household electricity is 27 cents per kW h. Therefore this air conditioner would cost approximately $10 to run for 5 hours.

5 Why is it that there are only two cables coming into the house from the street and yet power points always have three connections?

6 The function of a fuse is to burn out, and thus turn off the current, if the circuit is overloaded. Why is it always placed in the active wire at the meter box rather than the neutral one, given that this function could be fulfilled if it was in either?

7 What is the function of the 'earth stake' that will normally be found near a meter box?

8 A toaster cable with conductors coloured red, black and green is to be joined to another cable with brown, blue and green/yellow conductors. Peter has joined the red and blue, black and brown, and green and green/yellow. Will the toaster work normally when it is plugged in and turned on? Why is the way he has connected the cables dangerous?

9 An appliance was mistakenly wired between the active and earth instead of between the active and neutral. Explain why that is a very dangerous thing to do, even though the appliance will appear to work normally.

10 How much current would flow through a person with dry hands and a total contact resistance of 100 kΩ when they touch a 240 V live wire?

Chapter review

04

KEY TERMS

alternating current
circuit breaker
diode
direct current
earth
effective resistance
electric shock
fuse
junction
kilowatt hour
light-dependent resistor
light-emitting diode
overload
parallel circuit
potentiometer
residual current device
series circuit
short circuit
thermistor
transducer
voltage divider

1 Two resistors, R_1 and R_2, are wired in series. Which of the following gives the equivalent series resistance for these two resistors?

A $R_T = R_1 + R_2$
B $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2}$
C $R_T = R_1 - R_2$
D $R_T = R_1 \times R_2$

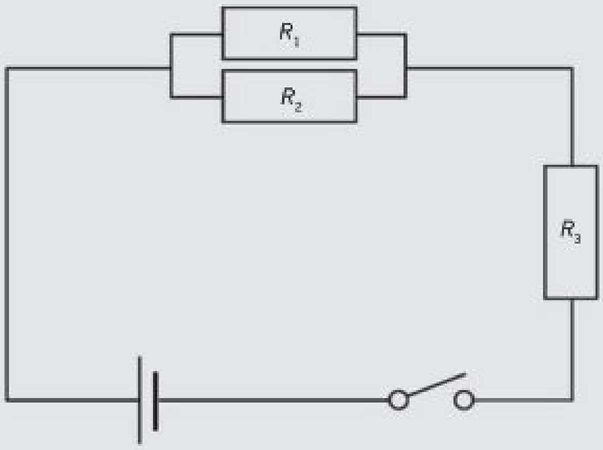
2 An electrical circuit is constructed as shown below. Use the information given about the circuit diagram to answer the following questions.

The electric cell provides 3 V.

The total resistance R_T of the circuit is 8.5 Ω.

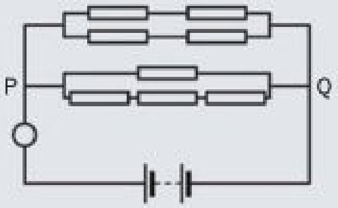
R_2 has a resistance of 15 Ω.

The total resistance of resistors R_1 and R_2 is 5.0 Ω.



a Find the value of R_3.
b Find the current through R_3.
c Find the potential difference across the parallel pair R_1 and R_2.
d Find the current through R_2.
e Find the current through R_1.
f Find the value of R_1.

3 Eight equal-value resistors are connected between points P and Q. The value of each of these resistors is 20.0 Ω.



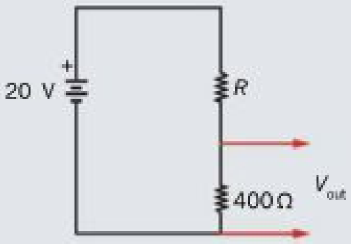
a The circle in the circuit diagram represents either an ammeter or a voltmeter. Identify which type of meter this should be.
b Calculate the total equivalent resistance of the circuit. Assume that the resistance of the meter and power source can be ignored.

4 Explain how an earth wire improves electrical safety in the home.

5 Sketch a circuit diagram showing how four 10 Ω resistors can be connected using a combination of series and parallel wiring to have a total equivalent resistance of 10 Ω.

6 A 9 V battery is connected to a voltage divider consisting of two resistors with values of 600 Ω and 1200 Ω. What would the voltage across the 600 Ω resistor be?

7 A voltage divider is constructed using a 20 V battery, a 400 Ω resistor and a variable resistor, R, as shown in the diagram. What should the resistance of R be in order to produce V_{out} = 8 V?



A 400 Ω
B 600 Ω
C 800 Ω
D 1200 Ω

8 Which of the following is an example of an LDR?

A a voltage divider
B an output transducer
C an input transducer
D a signal processing component

9 A circuit consists of a 12 V battery and three 20 Ω light bulbs. The bulbs are initially connected in series.

a Calculate the power output of the circuit.
b The circuit is changed so that the bulbs are connected in parallel. Calculate the power output of the circuit.
c Compare the power drawn in the parallel circuit with that of the series circuit.

10 Which of the following would be most likely to cause serious electrocution harm to a human being?

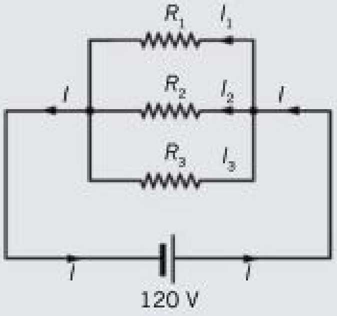
A high voltage spark from a Van de Graaff generator; duration = 1 ms
B 3 mA current; duration = 0.5 s
C 50 mA current; duration = 0.1 s
D 50 mA current; duration = 4.5 s

11 A 3 kW heating unit runs for 4 hours. If household electricity costs 30 cents per kW h, how much does it cost to run the heater for this time?

12 The best definition of power from the following choices is:

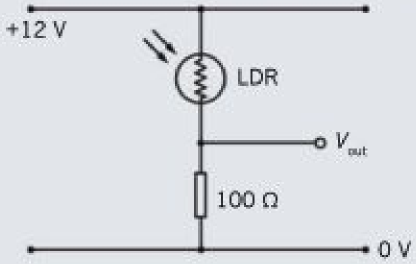
A the total amount of energy consumed by a circuit component
B the current drawn by a circuit component each second
C the rate at which voltage is supplied to a circuit component
D the rate at which energy is transformed by a circuit component.

13 Consider the following circuit where three resistors R_1, R_2 and R_3 are connected in parallel. Assume that $R_1 = 100\ \Omega$, $R_2 = 200\ \Omega$ and $R_3 = 600\ \Omega$. The battery has an EMF of 120 V.



a Calculate the total resistance, R_{total}, in the circuit.
b Calculate the line current, I, in the circuit.
c Determine the branch currents I_1, I_2 and I_3.
d What is the power output, P, of the battery?
e Calculate the total power consumed by all of the resistors in the external circuit.

14 In the simple LDR light-detector circuit shown below, V_{out} is to be used to activate an alarm when the ambient light reaches a certain level. At this particular light level, the resistance of the LDR is 200 Ω. The alarm activates whenever V_{out} is above the trigger level.



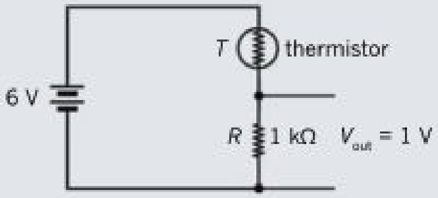
a What is the value of V_{out} at which the alarm should activate?
b Will the alarm activate when the light is above or below the particular level of concern? Explain your answer.
c When it is very dark, what would you expect V_{out} to become?

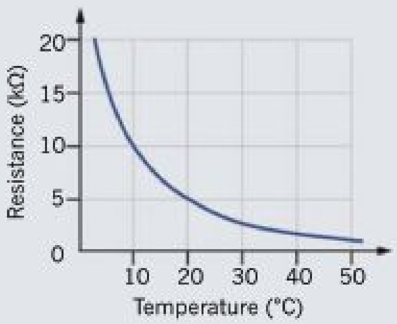
15 A student measures the resistance, R, of a thermistor for various temperatures, T, and records the results in the table below.

R (kΩ)	T (°C)	R (kΩ)	T (°C)
20.0	2.5	2.5	30
15.0	5.0	2.0	40
10.0	10.0	1.8	45
5.0	20.0	1.0	50

a What is a thermistor?
b Plot a graph of R versus T and describe whether the relationship between resistance and temperature is linear.

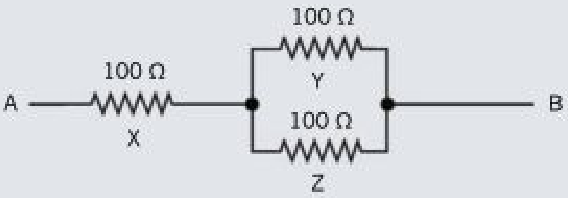
16 The graph below shows the resistance versus temperature characteristics for a particular thermistor. Determine the temperature of the thermistor in the circuit if V_{out} is 1 V.





Chapter review *continued*

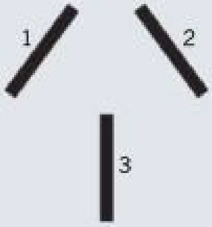
17 Three 100 Ω resistors are connected as shown. The maximum power that can safely be dissipated in any one resistor is 25 W.



a What is the maximum potential difference that can be applied between points A and B?

b What is the maximum power that can be dissipated in this circuit?

18 This diagram shows the three sockets when looking directly at a power point, numbered 1, 2 and 3.



a Which number is the earth socket?

b Which number is the active socket?

c Which number is the neutral socket?

19 Why is the shock received when a finger touches a live wire likely to be less severe than the shock received by a person who touches a live wire with a pair of uninsulated pliers?

20 It is said that a fuse protects property and a safety switch or RCD protects lives. Explain why this statement is true.

UNIT 1 • Area of Study 2

REVIEW QUESTIONS

How do electric circuits work?

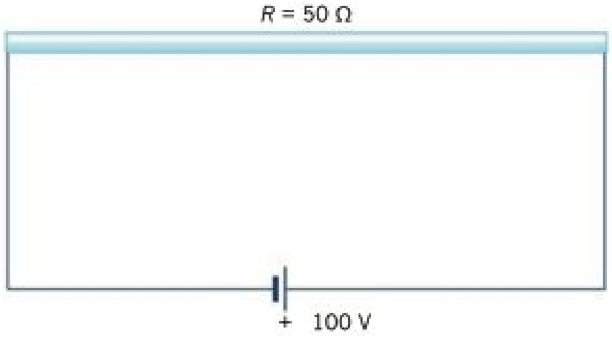
1 When a voltmeter is used to measure the potential difference across a circuit element it should be placed:

A in series with the circuit element
B in parallel with the circuit element
C either in series or in parallel with the circuit element
D neither in series nor in parallel with the circuit element

2 When an ammeter is used to measure the current through a circuit element it should be placed:

A in series with the circuit element
B in parallel with the circuit element
C either in series or in parallel with the circuit element
D neither in series nor in parallel with the circuit element

3 The following figure describes a simple electric circuit in which a length of resistance wire is connected to a battery ($q_e = 1.60 \times 10^{-19}$ C).



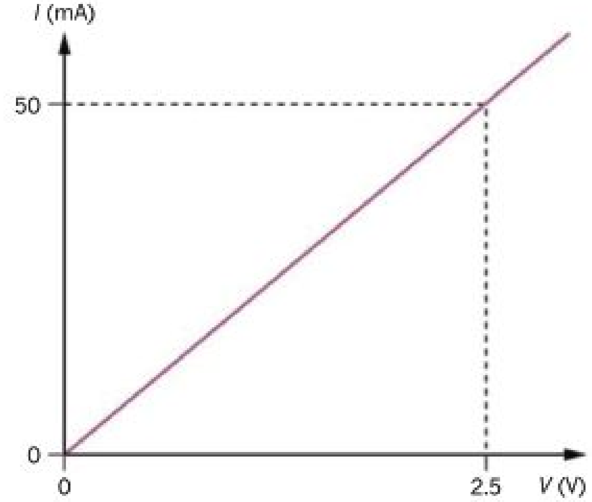
a How many electrons pass through the wire every second?
b How much electrical energy does each electron lose as it moves through the wire?
c What happens to the electrical energy of the electrons as they move through the wire?
d How much power is being dissipated in the resistance wire?
e What is the total energy being supplied to all the electrons passing through the wire each second?
f What is the power output of the battery?
g Discuss the significance of your answers to parts d and f.

4 Birds can sit unharmed on a high-voltage power line. However, on occasion large birds such as eagles are electrocuted when they make contact with more than one line. Explain.

5 Assume that when the dome of a Van de Graaff generator is fully charged it is at a potential of +400 kV and that the current from the belt onto the dome is a continuous 2.0 μA.

a Explain why the potential of the dome remains at 400 kV even though the belt continues to supply charge at the rate of 2.0 μA.
b How many electrons is the belt moving each second? Are these electrons being carried onto or off the dome?
c At what power must the motor be working to keep supplying the current to the fully charged dome?
d If the dome is at a potential of 400 kV, what is the potential energy (in J) of each elementary charge on it? ($q_e = 1.60\ \ 10^{-19}$ C)

6 The diagram below shows the current–voltage graph for a section of platinum wire. A potential difference of 9.0 V is established across the section of wire.



a Determine the resistance of the section of wire.
b How much energy does an electron lose in travelling through the wire?
c Calculate the power dissipated in the wire.
d How many electrons pass through the wire in 10 s?

7 When two circuit elements are placed in series:

A the power produced in both must be the same
B the current in both must be the same
C the voltage across both must be the same
D the resistance of the combination will double

8 When two circuit elements are placed in parallel:

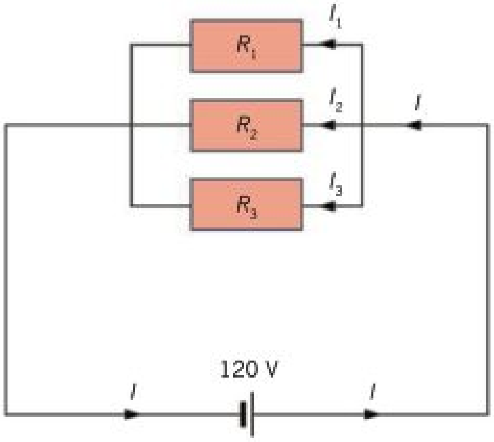
A the power produced in each element must be the same

B the current in each element must be the same

C the voltage across each element must be the same

D the resistance of the combination will halve

9 Three resistors, $R_1 = 100\ \Omega$, $R_2 = 200\ \Omega$ and $R_3 = 600\ \Omega$, in a circuit are connected in parallel with a 120 V battery.



a Calculate the total resistance in the circuit.

b Determine the total current in the circuit.

c Calculate the branch currents I_1, I_2 and I_3.

d Determine the power output of the battery.

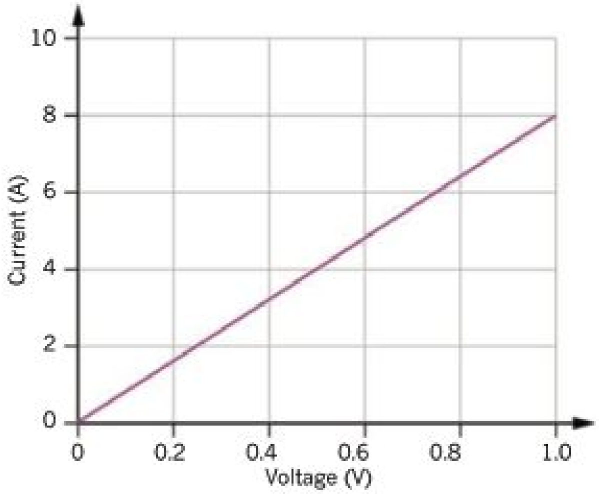
e Calculate the total power consumed by all the resistors in the circuit.

10 A student needs to construct a circuit in which there is a voltage drop of 5.0 V across a resistance combination, and a total current of 2.0 mA flowing through the combination. She has a 4.0 kΩ resistor that she wants to use and proposes to add another resistor in parallel with it.

a What value should she use for the second resistor?

b Determine the effective resistance of the combination.

11 An electrical extension cord used to run a 2400 W mains-powered heater is found to have the *I*–*V* characteristic shown in this graph.



a What is the resistance of the extension cord?

b How much voltage drop will occur in the cord while the heater is running?

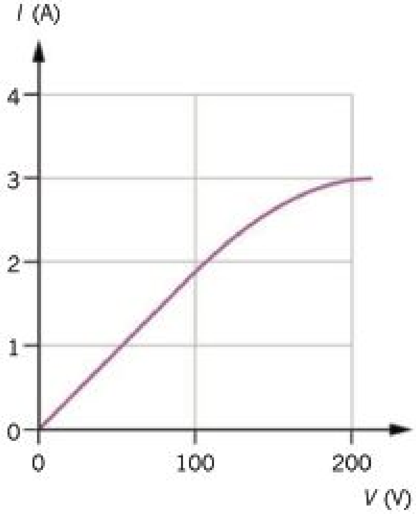
c How much power will be dissipated in the extension cord?

12 **a** In some ways the energy released in a light bulb in an electric circuit can be compared to the energy released as water falls down a waterfall. What power would be released by water falling 0.6 m down a garden waterfall at the rate of 20 L (20 kg) per min?

b What current would need to flow through a 3.0 V torch bulb to produce the same amount of power as the waterfall in part a?

The following information applies to questions 14 to 17.

A heater element is found to have the *I*–*V* characteristic shown in the graph below.



13 What current flows if a potential difference of 100 V is applied to the element?

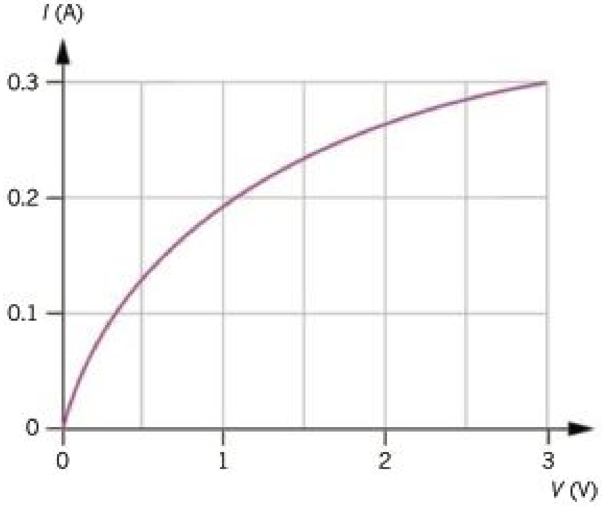
14 What is the resistance of the element at:

- **a** 50 V?
- **b** 100 V?
- **c** 150 V?
- **d** 200 V?

15 What power will be produced at 150 V?

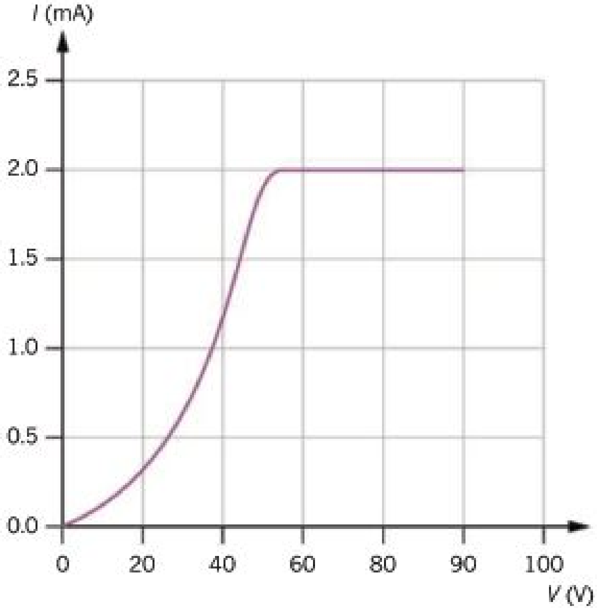
16 Why do you think that the resistance at 200 V would be greater than at 100 V?

17 The *I–V* characteristic of a small flashlight globe are shown in the graph below.

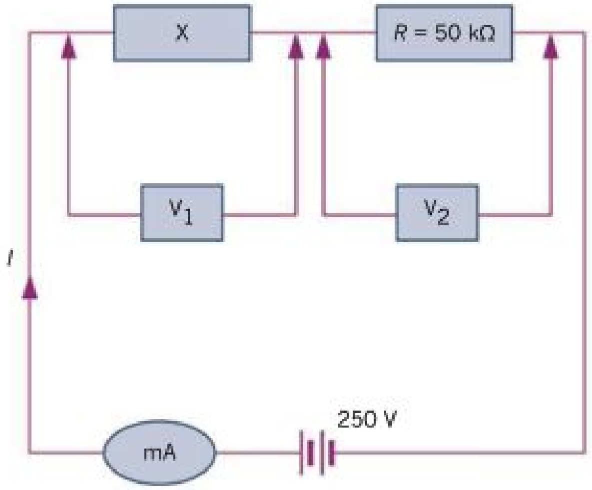


- **a** What current will flow in the globe when 2.0 V is applied to it? Would you expect twice the current to flow if the voltage is increased to 4.0 V? Explain.
- **b** Explain why this graph tells us that the globe is not an ohmic device.
- **c** How much power is the globe using when it is operating at 3.0 V?

18 A special type of circuit device X has the *I–V* characteristic depicted in the following graph.



When circuit device X is connected into the circuit, as shown in the circuit diagram below, the reading on the ammeter is 2.0 mA.



- **a** How would you describe the device X and what is its purpose?
- **b** Describe the resistance of such a device in the voltage range 60–100 V.
- **c** What are the readings on voltmeters V_1 and V_2?
- **d** How much power is being consumed by the circuit device X?
- **e** How much power is being consumed by the resistor?
- **f** What is the power output of the 250 V battery?

The following information applies to questions 21 and 22. A student performs an experiment in which an electric motor is used to lift a 200 g weight through 2.0 m, thus increasing its potential energy by 4.0 J. From measurements of the rate at which the weight is lifted the efficiency of the motor is to be determined. Two different voltages were used and the current was measured.

19 In the first experiment at 6.0 V, a current of 0.25 A was measured and the weight took 5.0 s to rise the 2.0 m. What was the efficiency of the motor?

20 In the next experiment the voltage was increased to 8.0 V. The current was found to be 0.30 A and the efficiency worked out to be 60%. How long did the motor take to lift the weight the 2.0 m this time?

The following information applies to questions 23 to 25. The 240 V supply cables to a certain house have a total resistance of 0.50 Ω. The maximum likely power use at any one time is estimated to be 10 kW, while the normal minimum load is estimated at 120 W.

21 What is the maximum current likely to be used by the household?

22 What will the voltage drop along the supply cables be under minimum load?

23 What will the voltage at the house switchboard be under maximum load?

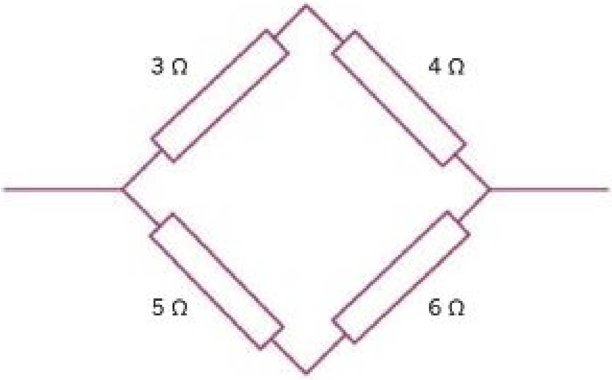
The following information applies to questions 26 to 28. Bill and Mary are discussing the lighting for their living room. At present they have four 60 W, 240 V light bulbs in parallel. Bill suggests that it might be cheaper to replace these with four bulbs wired in series.

24 If this was to be done, what would be the voltage and power rating of each of the new bulbs they would need in order to produce the same amount of power?

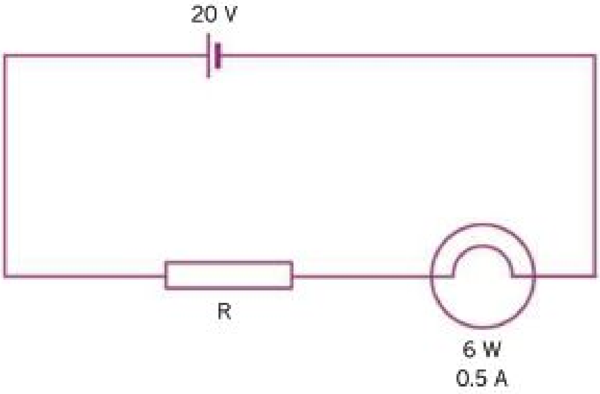
25 What would be the total current flowing in the circuit and how would this compare to the total current flowing when the original parallel bulbs were used?

26 Bill says that they would save on electricity bills because the current is going through all four bulbs and therefore being used more effectively. Mary says this is not right and that the power bill would be exactly the same. Who is correct and why?

27 What is the effective resistance of the four resistors shown in the diagram below?



28 A light bulb rated as 6.0 W and 0.50 A is to be operated from a 20 V supply using a limiting resistor, *R*. The circuit is represented below.

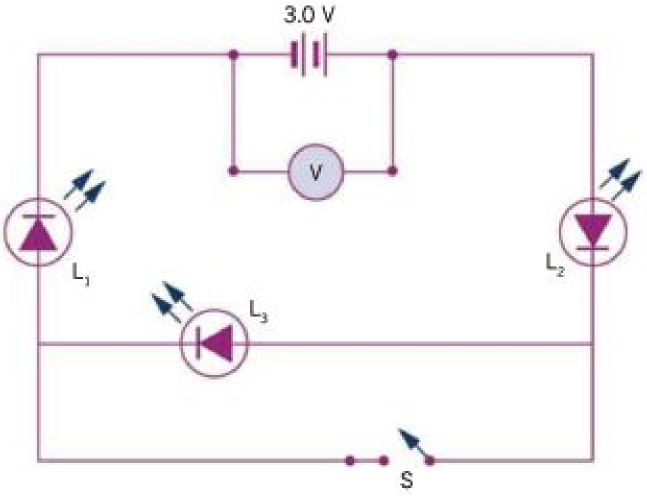


a What voltage is required to operate the light bulb correctly?

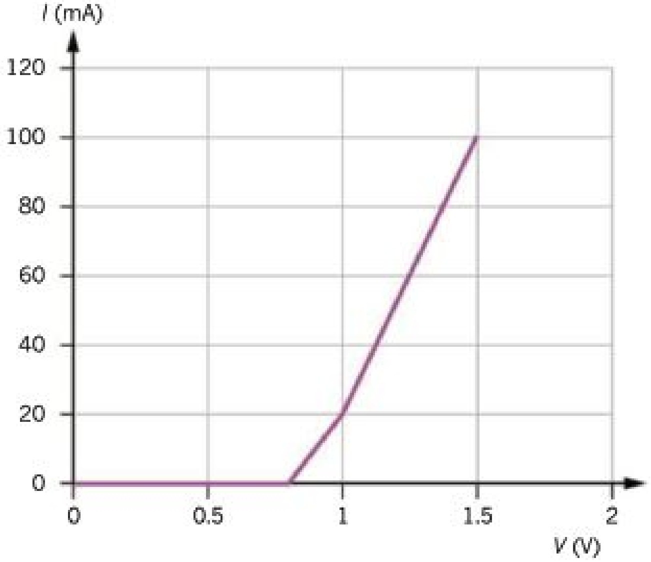
b How many volts will there be across *R* when the circuit is operating correctly?

c What is the required resistance of *R*?

The following information applies to questions 31 to 36. Three identical light-emitting diodes (LEDs) L_1, L_2 and L_3 are connected into the circuit as shown in the diagram below.

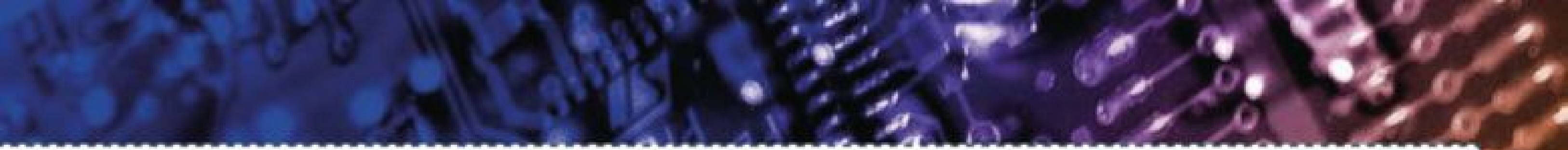


All LEDs are operating normally. The brightness of an LED increases when the current through it increases. The voltage–current characteristic for the LEDs is shown in the following graph. Assume that all connecting wires have negligible resistance.



29 When switch S is closed, the brightness of L_1 will:
A be greater than previously
B be less than previously but greater than zero
C remain the same
D be equal to zero

30 When switch S is closed, the brightness of L_3 will:
A be greater than previously
B be less than previously but greater than zero
C remain the same
D be equal to zero



31 When switch S is closed, the reading on voltmeter V will:

- **A** be greater than previously
- **B** be less than previously but greater than zero
- **C** remain the same
- **D** be equal to zero

32 When switch S is closed, the power output of the battery will:

- **A** be greater than previously
- **B** be less than previously but greater than zero
- **C** remain the same
- **D** be equal to zero

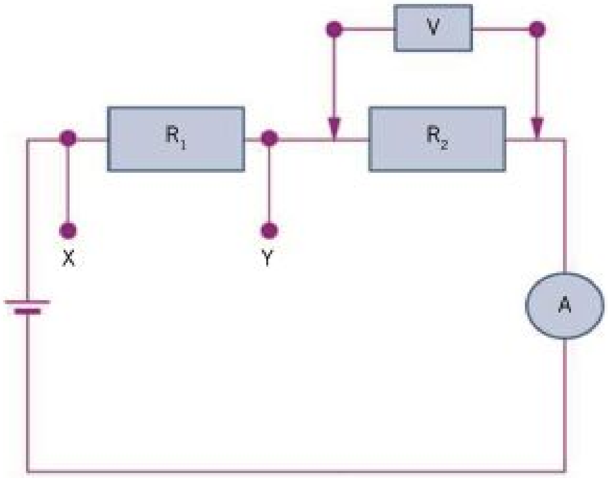
33 What is the effective resistance of an LED at a voltage of:

- **a** 1.0 V?
- **b** 1.5 V?

34 Determine the current in the circuit when:

- **a** the switch is open
- **b** the switch is closed.

The following information applies to questions 37 to 39. In the circuit shown, the current through $R_1 = I_1$ and the current through ammeter A = I_A. The voltmeter V has extremely high resistance and ammeter A has negligible resistance.



35 Which of the following statements is true?

- **A** $I_1 = I_A$
- **B** $I_1 > I_A$
- **C** $I_1R_1 = I_A(R_1 + R_2)$
- **D** $I_1R_1 = I_AR_2$

36 If another resistor R_3 was connected between X and Y, what would happen to the meter readings?

- **A** The reading on V would increase; the reading on A would decrease.
- **B** The reading on V would decrease; the reading on A would increase.
- **C** They would both decrease.
- **D** They would both increase.

37 If another resistor R_3 was connected between X and Y, what would happen to the power output of the battery?

- **A** It would increase.
- **B** It would decrease.
- **C** It would remain the same.

38 Explain the difference between a fuse and a residual current device in terms of their operation and the kind of protection that they provide.

The following information refers to questions 41 to 46. A household circuit is fitted with a 15.0 A circuit breaker. Susan connects the 800 W toaster, the 1200 W dishwasher and the 2000 W kettle in the same circuit.

39 Explain whether the devices in a household circuit are connected in series or in parallel and give a reason for this wiring.

40 Give a reason for the installation of the circuit breaker.

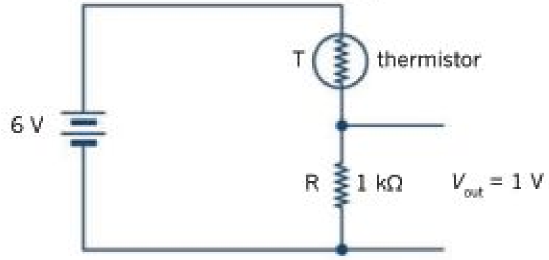
41 Calculate how much current the toaster draws.

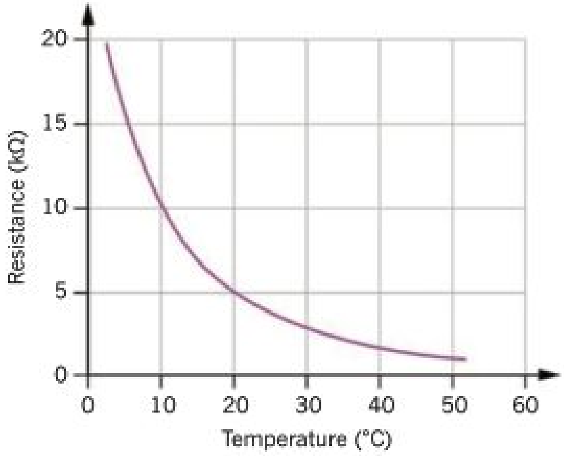
42 Calculate the resistance of the kettle.

43 Perform a calculation to predict whether or not the circuit breaker will trip the circuit when all three above devices are operating simultaneously.

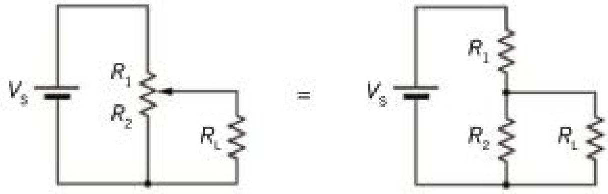
44 The dishwasher cycle is 2.5 hours. If Susan uses the dishwasher 6 times a week, what is the annual energy used in kW h?

45 A thermistor is a semiconductor device whose resistance depends on the temperature. The graph shows the resistance versus temperature characteristic for a particular thermistor. Determine the temperature of the thermistor in the circuit if V_{out} is 1 V.





46 Consider the potentiometer circuit shown below, and explain how it may be used to control the current through the load.



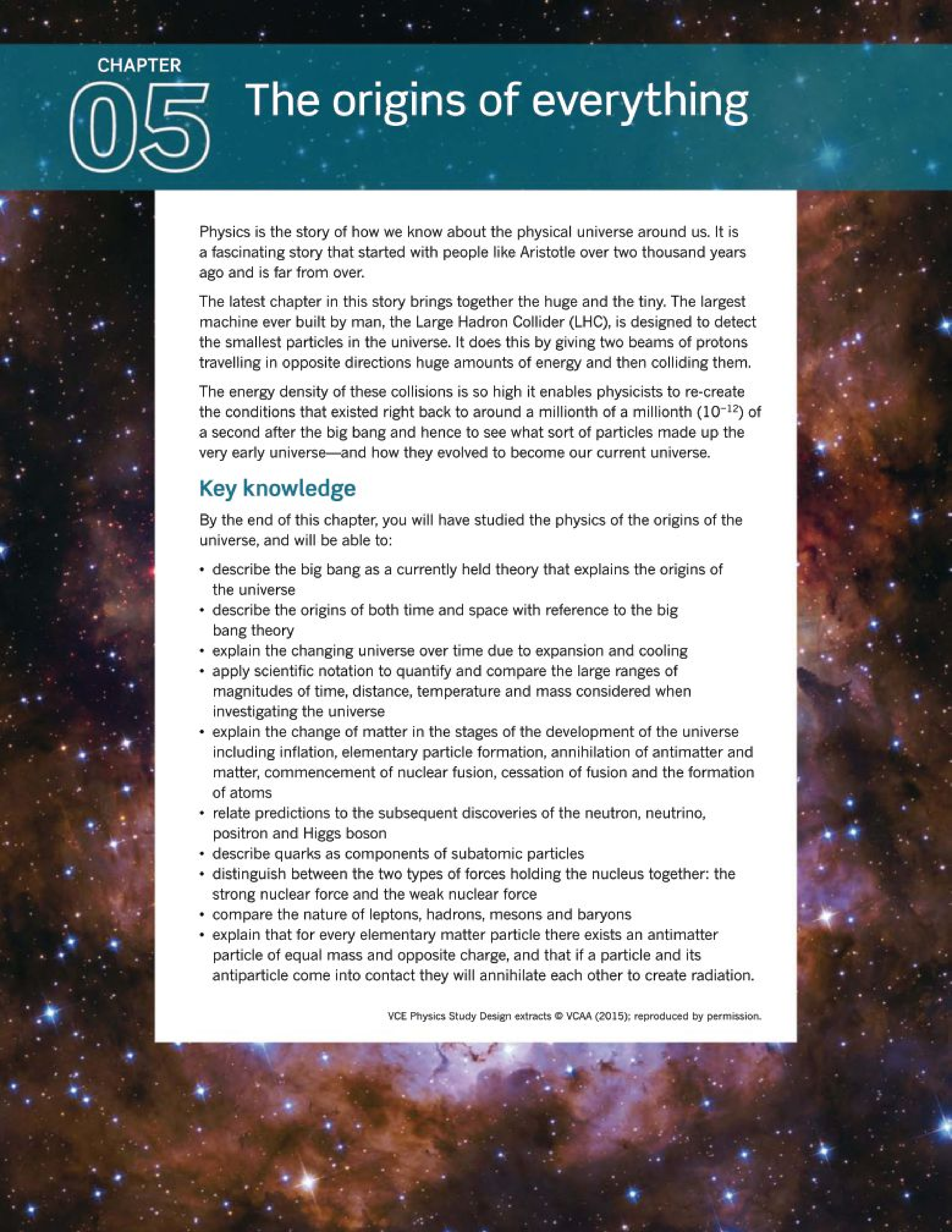
47 Explain why both a microphone and a speaker would be considered transducers, and describe the energy transformations occurring when each device operates.

48 When a person appears to have been electrocuted the rescuer should:

A pull them away from the electrical device
B call 000
C commence CPR immediately
D switch off the power first

49 Explain why a short circuit may trigger a circuit breaker.

50 Peter is puzzled because he sees that some household plugs have three prongs and some only two. He is concerned that the two-pronged plugs may be unsafe. Explain how this works to put his mind at rest.



CHAPTER 05

The origins of everything

Physics is the story of how we know about the physical universe around us. It is a fascinating story that started with people like Aristotle over two thousand years ago and is far from over.

The latest chapter in this story brings together the huge and the tiny. The largest machine ever built by man, the Large Hadron Collider (LHC), is designed to detect the smallest particles in the universe. It does this by giving two beams of protons travelling in opposite directions huge amounts of energy and then colliding them.

The energy density of these collisions is so high it enables physicists to re-create the conditions that existed right back to around a millionth of a millionth (10^{-12}) of a second after the big bang and hence to see what sort of particles made up the very early universe—and how they evolved to become our current universe.

Key knowledge

By the end of this chapter, you will have studied the physics of the origins of the universe, and will be able to:

- describe the big bang as a currently held theory that explains the origins of the universe
- describe the origins of both time and space with reference to the big bang theory
- explain the changing universe over time due to expansion and cooling
- apply scientific notation to quantify and compare the large ranges of magnitudes of time, distance, temperature and mass considered when investigating the universe
- explain the change of matter in the stages of the development of the universe including inflation, elementary particle formation, annihilation of antimatter and matter, commencement of nuclear fusion, cessation of fusion and the formation of atoms
- relate predictions to the subsequent discoveries of the neutron, neutrino, positron and Higgs boson
- describe quarks as components of subatomic particles
- distinguish between the two types of forces holding the nucleus together: the strong nuclear force and the weak nuclear force
- compare the nature of leptons, hadrons, mesons and baryons
- explain that for every elementary matter particle there exists an antimatter particle of equal mass and opposite charge, and that if a particle and its antiparticle come into contact they will annihilate each other to create radiation.

5.1 Measurements in the universe

For most of human history it was thought that the Earth was the centre of the universe. The Sun and stars were all just in the Earth's 'heavens'. Galileo realised that the Earth orbited the Sun and so thought the Sun was the centre of the universe, but it was Newton who realised that the Sun (see Figure 5.1.1) was in fact a star, like the thousands of others in the night sky. He said that our planet was one of six revolving around the Sun. He figured that if the other stars were similar to our Sun they must be something like a million times further away—way too far to measure by any means known at that time.

FIGURE 5.1.1 The brightest star in the sky—the Sun. It is a huge ball of very hot, very dense gas.

It was at this point that scientists' knowledge of the universe started to expand and develop into the picture that exists today. This includes our understanding of the vast distances involved and the amazing timescale over which the universe developed.

DEVELOPING A PICTURE OF THE UNIVERSE

With the development of better telescopes in the nineteenth century, it was noticed that some stars showed a very small movement against the background stars over the course of a year. It was realised that this was actually due to the Earth's motion around the Sun and the fact that these stars were closer than most of the other stars. The apparent motion is referred to as **parallax movement**.

PHYSICSFILE

Parallax

Hold your arm out straight and point towards the corner of the room, where the ceiling and two walls meet. Close one eye and line up your finger so it points exactly into the corner. Now open that eye and close the other without moving your finger. You will find that your finger is no longer pointing to the corner. If you keep your arm still, but keep swapping eyes, it will seem as though your finger is moving compared with the corner. This is an example of parallax movement.

As shown in Figure 5.1.2, the parallax angle, p, is clearly a measure of the distance to the star. The larger the angle, the closer the star. In fact, the key measure of distances in space is defined as the reciprocal of the parallax angle measured in arc seconds (that is, $\frac{1}{3600}$ of a degree). Stated simply, a star with a parallax angle of 1 arcsec is said to be at a distance of one **parsec** (1 pc).

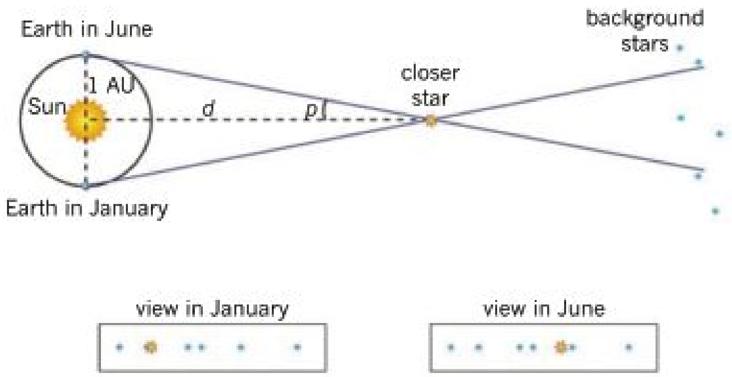


FIGURE 5.1.2 The parallax movement of a close star relative to the background 'fixed' stars. The parallax angle, p, is half the total apparent shift in angle.

$$\text{distance in parsec} = \frac{1}{\text{parallax angle in arcsec}}$$

$$d = \frac{1}{p}$$

PHYSICSFILE

Equating units of distance: AU and pc

Astronomers define the parallax angle as the angle subtended by the radius of the Earth's orbit (1 AU) as it would be seen at the distance of the star. This angle is labelled p in Figure 5.1.2. The total circumference of a circle of radius d would be $2\pi d$. As 1 arcsec is $\frac{1}{3600}$ of a degree, and a degree is $\frac{1}{360}$ of a full circle, we can see that 1 AU must be $\frac{1}{(3600 \times 360)}$ of the full circumference; that is:

$$1\text{ AU} = \frac{2\pi d}{(360 \times 3600)}$$

It can be shown that this leads to $d = 206\,265$ AU, which is indeed the conversion factor from parsec (pc) to astronomical units (AU).

Our closest star (apart from the Sun) is Proxima Centauri. It is a faint star close to Alpha Centauri, the 'Pointer' furthest from the Southern Cross. Proxima Centauri has a parallax angle of 0.77 arcsec, giving it a distance of 1.30 pc. Because of the huge distances in the universe, the kiloparsec (kpc, 10^3 pc) and megaparsec (Mpc, 10^6 pc) are also often used.

PHYSICSFILE

The light-year

A light-year is the distance that light travels in a year. Light travels 3×10^8 metres in one second. So it travels $3 \times 10^8 \times 60 \times 60 \times 24 \times 365.25 = 9.47 \times 10^{15}$ m in one year.

Two other measures of distances in space are often used:

- The distance from the Earth to the Sun is referred to as an astronomical unit (AU).
- A light-year (ly) is the distance light travels in a year.

The relationship between the three measures of distance are shown in Table 5.1.1. While the distance to the Sun can be thought of as huge, clearly it is tiny in astronomical terms!

	Distance in:			
	m	AU	Light-years	Parsec
1 AU	1.496×10^{11}	1	0.000016	0.000005
1 light-year	9.461×10^{15}	63240	1	0.3066
1 parsec	3.086×10^{16}	206265	3.2616	1

TABLE 5.1.1 Relationships between astronomical distance units.

It turned out that while the closest stars were around 300 000 times as far away as the Sun, most were well over a million times as far away. As telescopes became more powerful and we could see more of the universe, it was clear that the number of stars was, well, astronomical!

MEASUREMENT IN SPACE

In the study of the universe, the numbers for measurements are huge. You probably have a pretty good idea of what a metre looks like, but it is much harder to visualise a parsec.

To help you understand these numbers, consider the distances listed in Table 5.1.2.

	Approximate distance (m)
the radius of an electron	10^{-16} (approximate)
the radius of a carbon atom	7.0×10^{-11}
the size of a coffee bean	1.2×10^{-2}
the average height of a human	1.7
the height of Mount Everest	8.8×10^{3}
the distance from Melbourne to London	1.7×10^{7}
the distance to the Sun (from Earth)	1.5×10^{11}
the diameter of the Milky Way	1.0×10^{21}
the distance to the Andromeda galaxy	2.4×10^{22}
the radius of the observable universe	4.4×10^{26}

TABLE 5.1.2 Comparison of various-sized objects.

As you look at the origins of the universe and the creation of matter from subatomic particles, you'll see the timescale over which the universe was formed, the temperatures generated and the mass of subatomic particles. Like the comparisons of distances in Table 5.1.1 (p.169) , you'll find it easier to compare the measurements if everything is in the same unit and you use a unit with which you are familiar.

PHYSICSFILE

A billion

A billion is equal to 1×10^9, that is, 1 with 9 zeros after it. In other words, it's a thousand million. Originally the British version of a billion was a million million. The British unit was changed to be like the American unit, and now there is consistency around the world about the word 'billion'.

THE UNIVERSE IS HUGE

Today, scientists know that not only are there several hundred billion stars in our galaxy, but there are about a hundred billion galaxies. The nature and origin of all these stars and galaxies has puzzled scientists for many centuries.

FIGURE 5.1.3 This image shows many fuzzy objects in the night sky, now known to be galaxies.

At the start of last century, astronomers were curious about small fuzzy objects in the night sky called spiral **nebulae** (see Figure 5.1.3). There were many 'nebulae', but most were thought to be clouds of dust shining in the light of nearby stars. The spiral ones appeared to be different however.

It was Edwin Hubble, in the 1920s, who resolved the mystery. After careful examination of his photographs of spiral nebulae he discovered that some contained stars called 'Cepheid variables'. These were known to be very bright stars which slowly changed in brightness over a period ranging from a day or two to several weeks. It had been found that their **intrinsic brightness** (their actual, or real brightness regardless of their distance away) was related to the period of this variation.

EXTENSION

The brightness of stars

Astronomers measure the apparent brightness of stars on a scale that actually originated with Hipparchus in the second century BCE! He called the brightest stars he could see 'first-magnitude' stars (+1), those about half as bright 'second-magnitude' (+2) and so on to those barely visible, which were 'sixth-magnitude' (+6) .

When astronomers sailed into the southern hemisphere they discovered brighter stars and so the scale had to be extended 'upwards' to 0 magnitude and then –1 magnitude and so on.

When the telescope was invented the scale went 'downwards' to +7 and beyond.

The scale is referred to as the apparent magnitude scale. Don't be confused by the fact that the scale seems 'backwards'—dimmer stars have a numerically higher magnitude.

In the nineteenth century when astronomers were better able to measure the brightness of stars, they defined apparent magnitude more precisely by saying that a difference in magnitude of 5 corresponds exactly to a factor of 100 times in apparent brightness—which agreed roughly with the old values. In other words, it would take 100 stars of magnitude +6 to equal the brightness of one star of magnitude +1. Mathematically this means that each level of magnitude represents a change in brightness of about 2.5 times (instead of Hipparchus's 'double'). In other words, it would take 2.5 stars of magnitude +6 to equal the brightness of a single +5 star, for example. Or it would take 2.5 × 2.5 = 6.3 stars of magnitude +6 to equal the brightness of a +4 star. Figure 5.1.4 shows the scale with some examples. You might like to check with your calculator that it would take about 13 billion Sirius A stars to equal the apparent brightness of the Sun! (What we call 'Sirius' is actually two stars very close together, Sirius A and the much dimmer Sirius B.)

Clearly there must be a relationship between the apparent brightness of a star, its distance from Earth and its intrinsic brightness (actual brightness). Once the actual distances of a number of stars were known, it was possible to determine their intrinsic brightness. It was soon found that stars vary enormously in intrinsic brightness. For example, Sirius and Canopus are the two brightest stars in the sky; however, although Canopus appears almost as bright, it is actually about 36 times further away than Sirius. To make up for this big difference in distance, it can be calculated that it must be about 3000 times brighter than Sirius. The very closest star, Proxima Centauri, is not even visible to the naked eye and so must be a very dim star. In fact it is intrinsically only about $\frac{1}{30000}$ as bright as Sirius. The intrinsic brightness of stars varies over a huge range.

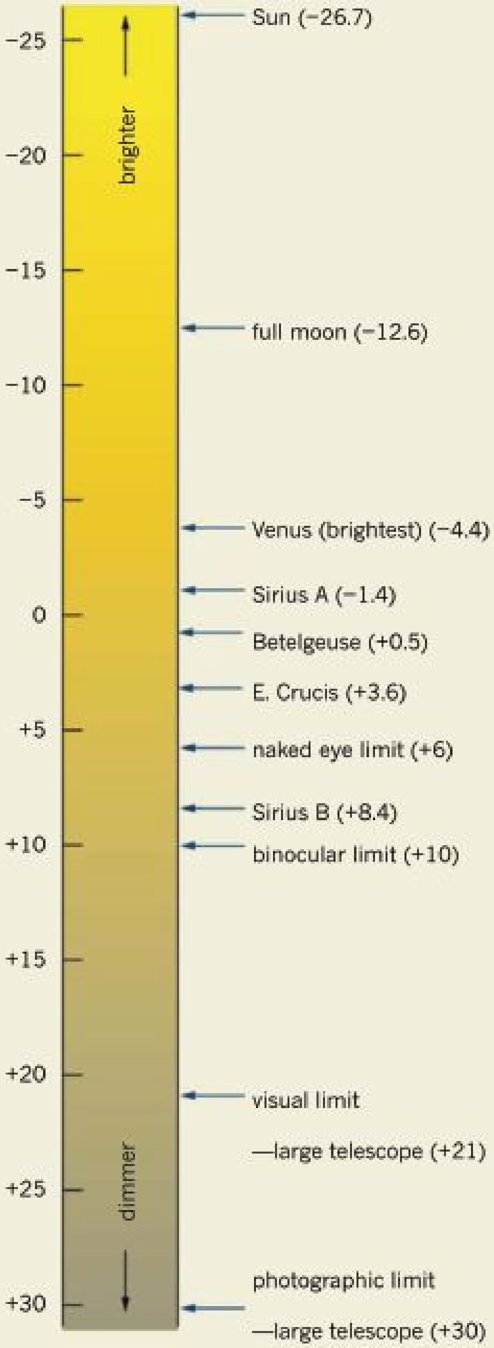


FIGURE 5.1.4 The apparent magnitude scale is a based on an ancient Greek scale with nineteenth-century mathematical corrections. Visible stars range from –1.44 (Sirius A) to about +6 for stars barely visible under the best conditions.

FIGURE 5.1.5 The Andromeda galaxy is one of hundreds of billions of galaxies in the universe. It is the closest to our own Milky Way galaxy.

It is possible to work out the distance to these stars by comparing their intrinsic brightness to their apparent brightness. If a star that is known to be intrinsically bright appears to be fairly faint, it is clearly a long way away. In this way Hubble found that these nebulae (galaxies) containing Cepheids were indeed a very long way away—well outside our Milky Way galaxy. The closest, the Andromeda galaxy, shown in Figure 5.1.5, is around 750 kpc away. Compare that with the diameter of the Milky Way, which is about 50 kpc.

Modern telescopes have been used to measure the distances of galaxies containing Cepheid variables out to about 30 Mpc away. The **Hubble Space Telescope**, shown in Figure 5.1.6, has helped astronomers to find galaxies around a thousand megaparsecs away. When they look at these galaxies, astronomers are effectively looking back in time because the light from them has taken so long to reach us.

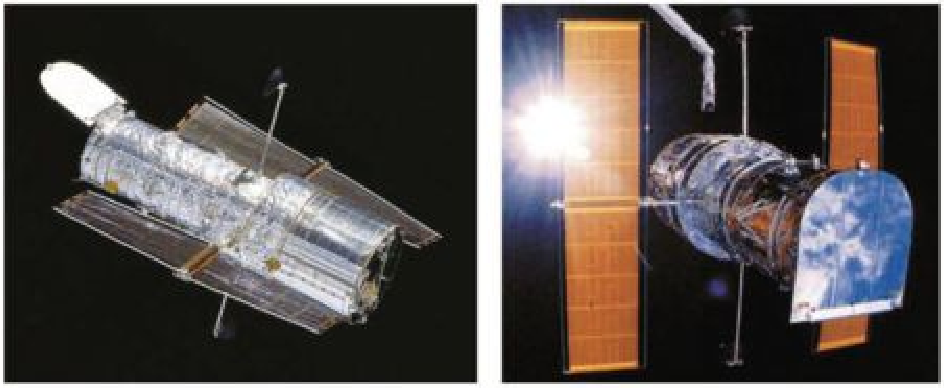


FIGURE 5.1.6 Probably the best-known telescope today is the Hubble Space Telescope. It weighs 13 tonnes and is 13.1 metres long, and has brought us magnificent pictures of astronomical sights.

Red shift and Hubble's law

Hubble did more than just show that the 'spiral nebulae' were actually galaxies like the Milky Way, but at huge distances away. When he carefully examined the light from stars in the galaxies he found that it showed the familiar spectra of the elements, such as hydrogen and helium, that we know on Earth, but all the lines were **redshifted**, in other words moved toward the red end of the spectrum (see Figure 5.1.7).

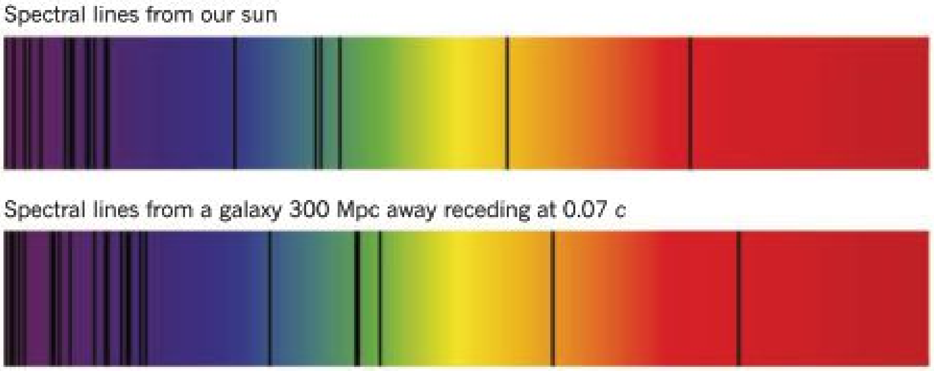


FIGURE 5.1.7 Notice that the black lines in both spectra form the same arrangement but the bottom set of lines are 'shifted' slightly to the right when compared to the lines in the spectrum from the Sun. This is known as 'redshift'.

He interpreted this redshift as meaning that the galaxies were rushing away from us at huge speeds. To see why he came to this conclusion, the way light travels needs to be considered along with what the 'redshift' actually is.

Light travels as waves and the wavelength of these waves determines its colour. Shorter wavelengths (around 5×10^{-7} m) are blue and longer ones (around 7×10^{-7} m) are red.

PHYSICSFILE

Spectra

The different colours of visible light correspond to different wavelengths (for more detail see Section 2.1, page 30). Imagine passing a whole visible spectrum of light through a gas. When light passes through the gas, it doesn't pass through unchanged. Different gases will absorb different wavelengths of the light. If the light that has passed through a gas is examined, any wavelength that was absorbed by the gas will be missing from the spectrum. This will appear as a black line corresponding to the absorbed wavelength. Examples of a spectrum missing some wavelengths are shown in Figure 5.1.7. Every gas has particular wavelengths it absorbs, so spectra like those in Figure 5.1.7 can tell us what gases the light passed through.

To understand the redshift, it is easier to think of sound waves first. Imagine an ambulance racing along the road emitting sound waves, as shown in Figure 5.1.8. As the ambulance speeds forwards, the sound waves in front are compressed (shortened). Those behind the ambulance are lengthened. Correspondingly the frequency in front is increased and the frequency behind is decreased. This effect is clearly heard when an ambulance passes us. The frequency, or 'pitch', of the siren drops as it passes. This effect is often referred to as the **Doppler effect**.

FIGURE 5.1.8 The Doppler effect. A moving source of sound produces a shorter wavelength and higher frequency sound in front, but a longer wavelength and lower frequency one behind.

A similar effect occurs with light waves, although the speed of the source is very much faster. The wavelength of the light from a source rushing away from Earth will be longer than that from a stationary source; that is, it will have moved toward the red end of the spectrum. This is the 'redshift' Hubble found. He assumed that the redshift for galaxies meant that they were speeding away from Earth.

At the time Hubble made this prediction (the early twentieth century) the universe seemed huge and infinite. The idea of galaxies all rushing away from Earth seemed very hard to accept. But not only were the galaxies rushing away, Hubble found that the further away the galaxy, the faster it seemed to be moving. When Hubble graphed his data (see Figure 5.1.9), he found that the 'speed of recession' was actually proportional to the distance away. This has become known as **Hubble's law** and is expressed by the simple equation:

$v = H_0 d$

where v is the speed of recession (in km s^{-1})

H_0 is Hubble's constant

d is the distance away (in Mpc).

H_0 is the **Hubble constant** and corresponds to the gradient of the graph. It is usually given in units of km s^{-1} per megaparsec. It has a value of around 70 km s^{-1} Mpc^{-1}. So, for example, a 'typical' galaxy at a distance of 100 Mpc would be rushing away from Earth at 7000 km s^{-1}. That's about the same as travelling from Perth to Sydney in half a second!

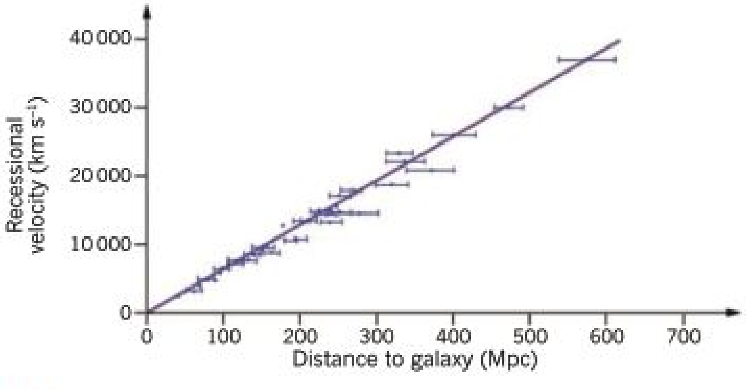


FIGURE 5.1.9 Hubble discovered that the galaxies were receding (rushing away) from Earth and that their recessional velocities were proportional to their distance from Earth.

Worked example 5.1.1

CALCULATING RECESSIONAL VELOCITY

Using a Hubble constant of 70 km s^{-1} Mpc^{-1}, calculate the recessional speed of a galaxy that is 450 Mpc from Earth.	
Thinking	**Working**
State Hubble's law.	$v = H_0d$
Substitute the given values into Hubble's law.	$v = 70 \times 450$
Calculate the speed.	$v = 31\,500$ km s^{-1}

Worked example: Try yourself 5.1.1

CALCULATING RECESSIONAL VELOCITY

Using a Hubble constant of 70 km s^{-1} Mpc^{-1}, calculate the recessional speed of a galaxy that is 250 Mpc from Earth.

EXPANSION OF THE UNIVERSE

Any relationship between the movement of the galaxies and the Earth's position in the centre of the universe seemed highly unlikely to the astronomers of the early twentieth century. The universe seemed to be basically uniform and infinite, and therefore seemed to have no centre. Also, the idea that there could be some sort of 'edge' to the universe didn't make sense.

Even Einstein, when he produced his general theory of relativity in 1915, was dismayed to find that it did not allow for an infinite universe. He amended the theory by adding what he called a 'cosmological constant' so that an infinite universe could exist. (Later he said this was his 'greatest blunder' because if he had believed his own equations it would have been he who discovered the expanding universe.)

Hubble explains expansion

Hubble eventually proposed that because all the distant galaxies were receding from Earth, space itself was expanding rather than the galaxies rushing away from us through space. This is just what Einstein's equations had predicted, but even he had not wanted to believe such an outrageous idea.

Hubble suggested that at a very large distance from the Earth, space will be expanding faster than the speed of light. Therefore the light from these stars would never reach us. Although this distance is not the 'edge of the universe', it is a distance beyond which it is impossible to ever see—the 'edge of the visible universe'.

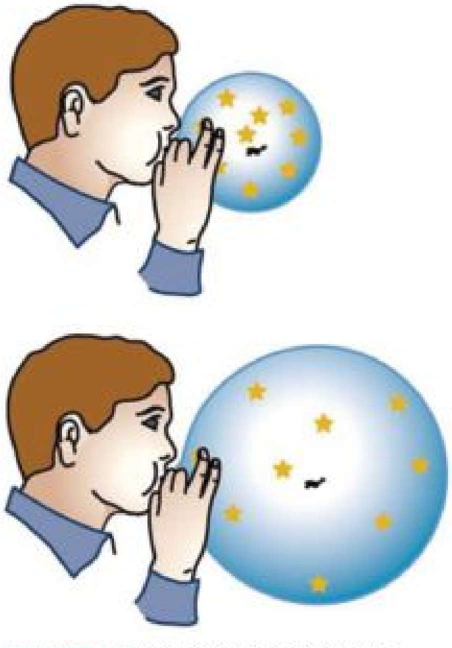


FIGURE 5.1.10 As the balloon is blown up the stars all move away from the ant. The more distant stars will move away faster than the closer ones. Remember that this is a two-dimensional analogy for what is really a three-dimensional universe.

Modelling the expanding universe

The idea that space itself is expanding is complicated. It is easy enough think of things expanding *in* space—a balloon being blown up, for example. Space seems so permanent and fixed, but this is not the case. It can be helpful to think of a two-dimensional analogy of what is really a three-dimensional problem. Imagine a little ant on the surface of that balloon being blown up, as shown in Figure 5.1.10.

The balloon is so big the ant thinks of the surface as flat (as we do of the Earth's surface). The balloon has little paper stars stuck onto it. As the balloon expands the ant would see all the stars moving away from it. Wherever it wandered on the balloon's surface the same pattern would apply—and the further away the stars the faster they would be receding.

The stars on the balloon were stuck on and not drawn on for a good reason. As the balloon expands, drawn stars would expand also, but the stick-on stars stay the same size. This is how astrophysicists picture the expansion of space. It is space that is expanding, not the stars and galaxies in it.

Expansion versus explosion

It is very important to make the distinction between this expansion of space and a conventional explosion, in which everything flies out from one point in space. To see uniform expansion in the explosion we would have to be right at the centre. This is not the case for the expansion of the universe (or the ant on the balloon). Astrophysicists talk of this as the 'cosmological principle', which states that the universe will appear uniform in whatever direction we look and from wherever we look. There is no 'centre' of the three-dimensional space of the universe.

PHYSICS IN ACTION

The future of the universe

The question about what will eventually happen to the universe has always fascinated astrophysicists. If a rocket is fired upwards it may reach a maximum height and fall back, or it may have enough energy to escape from the Earth and sail on into space forever. Or it may have just enough energy to escape but none left over for space adventures.

The question the astrophysicists asked was 'Did the universe have enough energy to overcome the gravitational attraction of all its mass and expand forever?'. Or would gravity eventually win out and pull it all back into a 'big crunch'—and perhaps start the cycle all over again with another big bang? Actually, because of Einstein's theory of general relativity, which links space, matter and gravity, the question was put a little differently. Mass, Einstein said, distorts space. We see the effects of the 'curved space' as the effects of gravity.

In the language of relativity, the question becomes whether space has positive, negative or zero curvature.

Positive curvature implies that there is enough matter in the universe to fold it in on itself—producing a closed universe that will eventually collapse to the big crunch. Negative curvature is the opposite—space curves outwards in an open-ended way and the expansion can continue forever. In between positive and negative curvature there is the 'flat' universe—equivalent to the expansion just being sufficient to go on forever, but only just. These three possibilities are sometimes represented by the two-dimensional analogies shown in Figure 5.1.11.

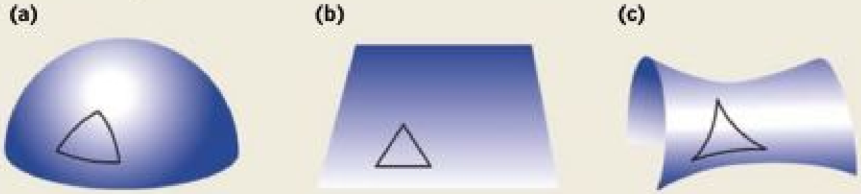


FIGURE 5.1.11 These diagrams are two-dimensional analogies for what is really three-dimensional space—or four, if time is included. In (a), positive curvature, the surface eventually folds in on itself, while the negative curvature of (c) can go on expanding forever. Surface (b) is the in-between 'flat' case.

There are different approaches to answering this question. One is to try to measure all the mass in the universe and see if it is sufficient to curve space 'inwards'. The difficulty here is to detect the mass. Most of the mass of the universe appears to be so-called 'dark matter', which by its nature is very hard to detect. Another approach is to measure the rate of expansion at different times in the history of the universe to see if it is slowing down. This can be done by looking at galaxies that are vast distances from us—the further away, the further back in time we are looking.

It is the cosmic microwave background radiation (CMB) that has given us our best clue to the answer. The CMB has been travelling through space since the time of the hot early universe. So, it has been affected by the space it has travelled through. It is possible to simulate the effects of positive, zero or negative space curvature on the radiation and then compare these with the actual radiation detected by satellites. The results of this research seem fairly clear—the CMB pattern is consistent with a flat universe: that is, zero curvature.

Given that the universe is flat, the total amount of mass and energy in it can be calculated. It turns out that even with the estimated amount of dark matter, there is still a huge amount of missing 'mass–energy'. Visible matter accounts for about 4% and dark matter for about 26%. This would appear to leave 70% of the total mass–energy content of the universe missing. However, astrophysicists now think they know where it is!

Perhaps the most surprising results have come from the studies of very distant galaxies. These studies suggest that rather than slowing down, the rate of expansion of the universe has been accelerating. At first this seemed to be quite contrary to all expectations. Even our spacecraft with plenty of energy to escape the Earth will still slow down as gravity tugs on them. Similarly, it was thought that the mass of the universe would eventually slow its expansion rate. However, when this acceleration was put together with the picture of a flat universe, as well as the implications of Einstein's 'cosmological constant', a new idea emerged. It seemed that the missing 70% of the mass–energy content of the universe might be in the form of what is called 'dark energy'.

This energy, it is thought, is what is actually accelerating the expansion—it is a sort of 'antigravity'. As the effect of gravity (space curvature) becomes less with the increasing size of the universe, the effect of the dark energy becomes more significant and the acceleration increases. A comparison of the different futures of the universe is shown in Figure 5.1.12, where graph (c) models the size of the universe according to the dark-energy model.

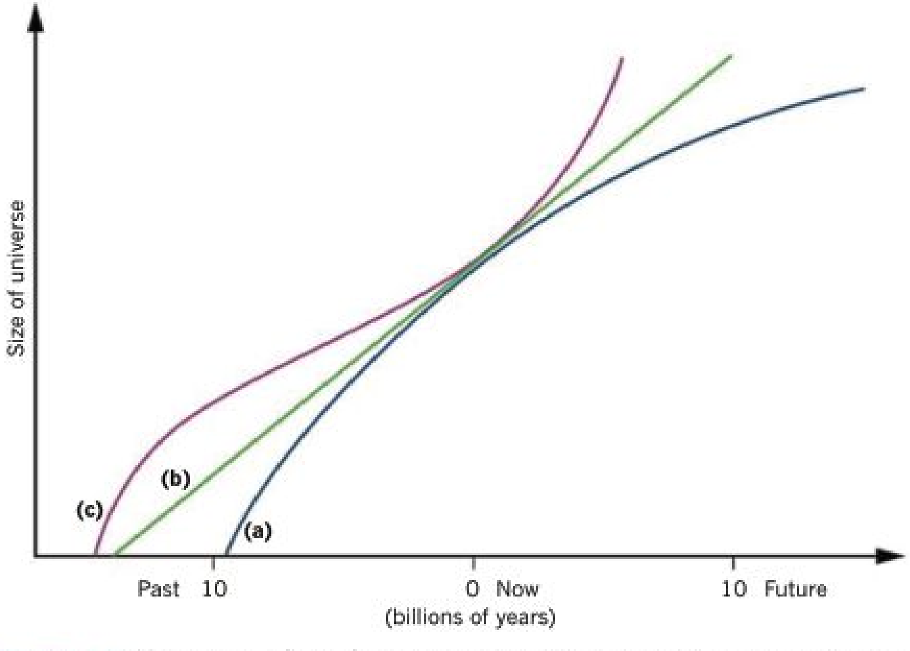


FIGURE 5.1.12 (a) This curve is for the flat universe without dark energy, and until recently seemed the most likely model. (b) This curve shows the expansion if we simply assume the Hubble constant has been constant from the beginning. (c) This curve is the latest model, which takes account of dark energy.

So, it does not seem that the universe will end with a big crunch after all. That would at least have given rise to another big bang! Rather, as everything accelerates apart it will get thinner and thinner until everything is so far away that it will become a very boring place with a very dim and uninteresting night sky.

5.1 Review

SUMMARY

- Distances in the universe are huge. The closest stars are a few parsecs away, but galaxies are millions of parsecs away.
- Most galaxies appeared to be 'redshifted'. This indicates that they are moving away from us. The further away, the faster this 'recessional velocity'.
- Recessional velocity can be found using Hubble's law:
  $v = H_0d$
- As the universe appeared to be uniform and with no centre, the redshift was eventually interpreted not as a Doppler shift of galaxies moving through space, but as the result of space itself expanding.

KEY QUESTIONS

1 Sirius, the brightest star in our sky, is 2.6 pc away from us. What is the change in its parallax angle in arcseconds? Give your answer to two decimal places.

2 Proxima Centauri is 1.30 pc from Earth. What is the distance to Proxima Centauri in metres? Give your answer to two significant figures.

3 What is the distance to Proxima Centauri in light-years? Give your answer to two significant figures.

4 What is the distance to Proxima Centauri in Astronomical Units? Give your answer to two significant figures.

5 Find the distance in metres to a galaxy 1000 Mpc away.

6 Use Figure 5.1.9 on page 173 to estimate the value of Hubble's constant H_0 in km s^{-1} Mpc^{-1} to two significant figures.

7 Use Hubble's law to determine the distance to a galaxy whose velocity of recession is 3.6×10^7 m s^{-1} and with an accepted value for H_0 of 72 km s^{-1} Mpc^{-1}.

8 Which one of the following is implied by the observation that all galaxies are moving away from us?

A. We happen to be at the centre of the universe.
B. We are at the place where the explosion called the big bang occurred.
C. The space between galaxies is expanding.
D. The galaxies are still rushing away from the place where the big bang was.

9 Not only did Hubble find a way to determine the distance to far-off galaxies, but he discovered that the further away they are, the faster they appeared to be moving away from us. How did he determine this?

10 A galaxy is determined to be approximately 350 Mpc from us. Use a Hubble constant of 70 km s^{-1} Mpc^{-1} to determine its recessional speed.

5.2 The big bang

Throughout recorded history, scientists have always sought to find an answer to how the universe began. They have also wondered if, when and how it will end. Astronomers have constructed hypotheses called cosmological models to try and find the answer. In the 1940s, the steady state theory and the big bang theory were the two contenders.

The big bang theorists won out in the end, and it's now the main theory explaining the origins of the universe. This section looks at the big bang theory (Figure 5.2.1) and how the evidence supports this theory.

FIGURE 5.2.1 The big bang is often thought of as an explosion, but it's much more complicated than that.

HOW THE UNIVERSE BEGAN

Two competing theories

Logically, if space is expanding, then at some time in the past it must have been just a dot. However, in the 1950s astronomer Fred Hoyle didn't agree, and he put forward what became known as the **steady state theory**.

He suggested that the universe was:

- infinite—the 'outer' stars would never reach infinity and so could go on moving away from us forever.
- expanding—matter is being created all the time at just the right rate to keep the density of the universe constant.

This was not really such an outrageous idea. **Quantum mechanics** had already suggested that matter was less permanent than originally thought. And the alternative was that everything was created in one 'big bang'. This is possibly harder to believe.

The **big bang theory** was the main competitor to Hoyle's steady state theory. It stated that galaxies outside the Milky Way were moving away from us. This implied that if time is run backwards, at some instant in time, the universe and everything in it was contained in a single point. In other words, everything was created from nothing in one 'big bang'.

It was in fact Hoyle who first used the expression 'big bang' to describe a universe that started off from a tiny dot. At the time, he was intending to be derogatory, but the name stuck.

Searching for the evidence

The astrophysicists of the day really were caught in a dilemma. Either way they had to accept the unacceptable: that matter just came into being out of nothing—either continuously or in one 'big bang'.

It seemed impossible to resolve this dilemma. If Hoyle was right, then the new matter being created should be able to be found somewhere. And as for the big bang, the only observational evidence that the universe was expanding, was consistent with either theory.

If matter was being created at a steady rate it was not too difficult to calculate that about 2 or 3 atoms of hydrogen would need to be created every day in a volume about the size of a large sports arena to account for the observed density and expansion. That didn't seem too impossible, but there was not much point in trying to look for it. The universe is so big, it would be hard to know where to look.

If the big bang theory was true, the universe would have been denser at an earlier time. It seemed that only some means of looking back in time could resolve the problem.

PHYSICSFILE

Time, space, matter and...

It is important to understand that the big bang was not seen as some sort of explosion from a small point in space. It was more an explosion of space, or more correctly an explosion of space–time.

Einstein had already showed that time and space were not separate entities. So the big bang would not have occurred at some point in time any more than it would have occurred at a point in space. Time, space and matter were all created together in the one 'big bang'.

EVIDENCE FOR THE BIG BANG

In the 1960s physicists at Princeton University were trying to resolve the question of the steady state theory versus the big bang theory. They decided that if the big bang did in fact happen, it must have been extraordinarily hot at the beginning. If so, it would have given off powerful heat radiation, as does any hot object. (More detail about how objects radiate heat can be found in Chapter 2.)

They also proposed that as the universe expanded, that radiation would have expanded as well. They calculated that the wavelength would have increased to around a millimetre by now (as shown in Figure 5.2.2). That means that, in effect, the apparent temperature of the radiation would have fallen from billions of degrees to just a few degrees above absolute zero.

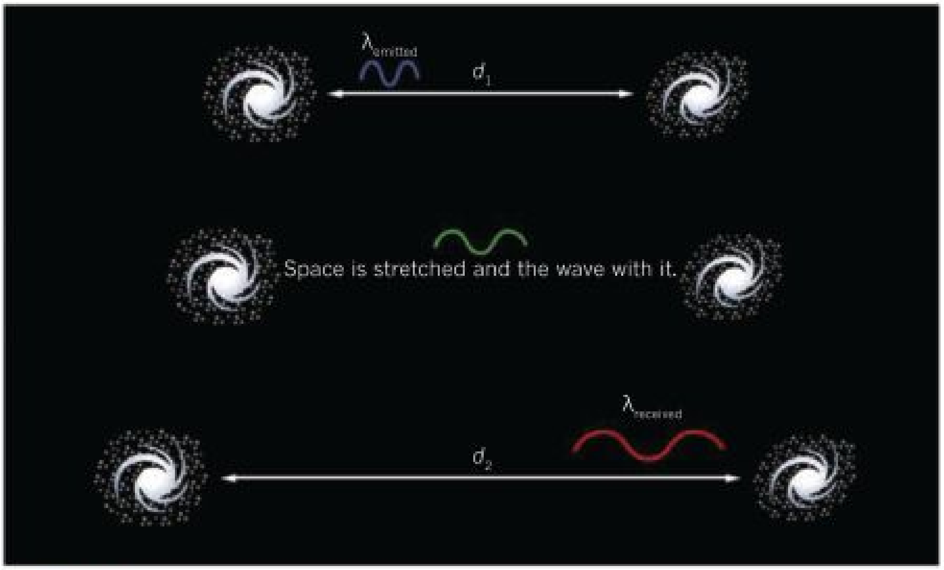


FIGURE 5.2.2 As the universe expanded, so too would the wavelength of the heat radiation left over from the big bang.

As it happened, two other physicists, Robert Wilson and Arno Penzias, at the nearby Bell Telephone labs, had discovered that their new antennas, designed to pick up radio signals from satellites, seemed to be picking up 'interference' from space. It turned out that this radiation was at just the sort of wavelength that Princeton physicists were looking for. Furthermore, it had the characteristics of heat radiation of just 2.7 K (that is, 2.7 degrees above absolute zero) as predicted. This was solid evidence in favour of the big bang theory at last, as there was no mechanism by which the steady state theory could produce this sort of radiation.

PHYSICSFILE

The source of interference

Wilson and Penzias were not sure at first where the interference picked up by the antenna was coming from. They thought it could have come from the Sun, but then ruled this out after some investigation. They made sure the interference wasn't from sources in the cities. They even ruled out any effect from pigeons by removing them and their droppings from around the antenna. Their realisation of what actually caused the interference was unexpected and absolutely ground breaking.

This left-over radiation came to be called the cosmic microwave background radiation, or CMB for short. Arno Penzias and Robert Wilson won the Nobel Prize for this discovery. Special satellites (see Figure 5.2.3) have mapped it carefully, as it is, in effect, a radio fingerprint of the early universe.

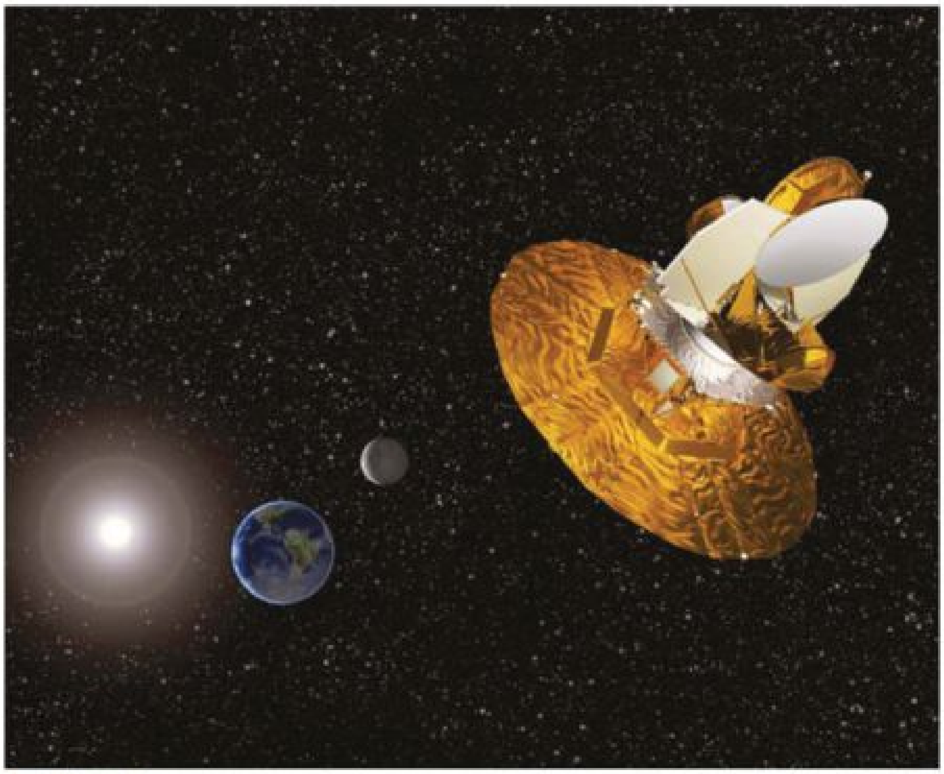


FIGURE 5.2.3 The WMAP satellite mapped the cosmic microwave background radiation and discovered tiny fluctuations in it, which probably marked the beginning of the formation of galaxies.

The CMB radiation is very uniform from all directions in the universe. This is to be expected, because it filled the early universe and still does today. However, it is not totally uniform. The tiny variations in the CMB radiation represent differences in temperature of only about 0.0003 K, but they are very significant because it means the early universe was not totally uniform. Had the early universe been uniform, it would not have been able to create stars and galaxies. The small variations meant that matter was slightly clumped together and so its unbalanced gravity could gradually pull it together to form the stars and, on a larger scale, galaxies.

THE BIG BANG

The universe seems to be simply one out of a virtually infinite number of possible universes that happened to have just the right conditions to enable life to emerge. A remarkably good picture of the first fractions of a second of the universe after the bang has emerged. It is a picture created from the known laws of physics, the study of interactions of very-high-energy particles in machines such as the Large Hadron Collider at CERN and a lot of hard work and imagination.

PHYSICS IN ACTION

How old is the universe?

Given that the universe started with a big bang, it should be possible to determine when it all started. In other words, it should be possible to find out how old the universe is. If Hubble's law is right, all we need to do is run the rate of expansion backwards to see when all the galaxies were in the same place. This is not so difficult.

Hubble's constant, H_0, is the proportionality constant between the distance from us to a galaxy and the galaxy's recessional velocity ($v = H_0 d$). If the galaxy has been receding from us at a speed v, and in that time has travelled a distance d, the time it has taken is $T = d/v$. Now the value of d/v from Hubble's equation is just $1/H_0$, so $T = 1/H_0$. Notice that the distance, d, is cancelled out in obtaining this expression—all galaxies, no matter how far away they are now, must have started off at the same point in time, i.e. the beginning of the universe. This is when the big bang happened. The age of the universe then is simply the reciprocal of Hubble's constant. Of course we need to convert the units into ones we recognise, as follows.

The age of the universe is the reciprocal of the constant:

$$T = \frac{1}{H_0}$$

$$= \frac{1}{70}\ \text{Mpc km}^{-1}\ \text{s}$$

The expression for the constant includes two different units for distance, km and Mpc, so we need to convert Mpc to km.

$1\ \text{pc} = 3.086 \times 10^{16}\ \text{m}$

This means that $1\ \text{Mpc} = 3.1 \times 10^{19}\ \text{km}$ (to two significant figures) and so:

$$T = \frac{3.1 \times 10^{19}}{70}$$

$$= 4.4 \times 10^{17}\ \text{seconds}$$

This is equal to:

$$T = \frac{4.4 \times 10^{17}}{(365 \times 24 \times 3600)}$$

$$= 1.4 \times 10^{10}\ \text{years}$$

That is, 14 billion years.

In fact, the shortcut calculation for converting from the Hubble constant to the age of the universe in billions of years is to use $980 \div H_0$. Hubble originally found a value for H_0 of 530, which would have given the age of the universe as $980 \div 530 = 1.85$ billion years.

This method of determining the age has assumed a constant rate of expansion and has ignored the effects of gravity. Neither of these are totally valid assumptions, but astrophysicists now have more complex models that do take these effects into account. Curiously enough, these other effects tend to cancel each other out and the models still produce an age close to 14 billion years. In fact the current best estimate is 13.8 billion years.

Inflation

In order to explain the initial stages of the big bang, astrophysicists found that there must have been an incredible period of 'inflation' right at the beginning. This period only lasted about 10^{-24} seconds, but during that time the size of the universe expanded to about 10^{50} times its original size.

This short period of inflation was necessary in order to prevent a very rapid collapse of the initial universe back into a black hole. Had the inflation lasted any longer, however, the greater expansion would have meant that atoms would have never formed.

Matter and antimatter

To understand where the matter in the early universe came from, a brief introduction to the Heisenberg uncertainty principle is needed. It says that there is a fundamental uncertainty in the position and momentum of any object. The more accurately one variable is known, the less accurately we can know the other variable.

It can also be interpreted as an uncertainty between mass and time. So in any very short time there is a fundamental uncertainty about mass. This is proportional: the shorter the time, the greater the uncertainty. The actual meaning of this is that in a very short time, mass can come in and out of existence from nothing.

Furthermore, when particles are created, they always come in pairs. For every matter particle there is an **antimatter** particle. But because they are so close when they pop into existence, they immediately **annihilate** each other (convert to energy) as can be seen in Figure 5.2.4. So matter (and antimatter) are actually popping in and out of existence everywhere all the time. You can't see it because it happens so quickly.

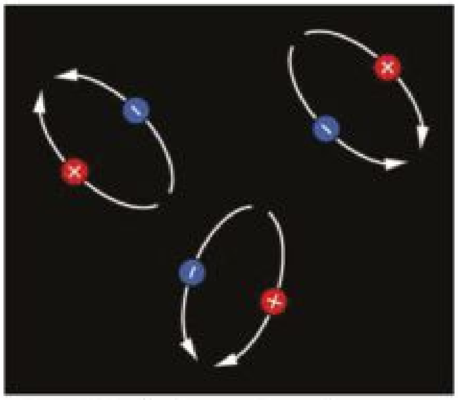


FIGURE 5.2.4 Electron–positron pairs were rapidly going in and out of existence in the early universe, but as they met they annihilated.

Creation of matter

To return to the origin of our universe, the two ideas just discussed can be combined—the inflation at the very beginning, and the continual creation and annihilation of matter. As two opposite particles are always produced in this creation, the process is called **pair production**. Due to annihilation, normally pair production doesn't result in the creation of any lasting matter. However, in the period of inflation, because of the extremely rapid expansion of space, pairs of particles rapidly became separated and didn't get a chance to annihilate. And so in that tiny fraction of a second of inflation, huge amounts of matter were created.

In those first moments after inflation, the universe was absolutely chaotic, with particles and antiparticles annihilating each other and producing high-energy gamma photons (light). Those photons themselves again collided with others and their energy formed new particle pairs. This is illustrated in Figure 5.2.5.

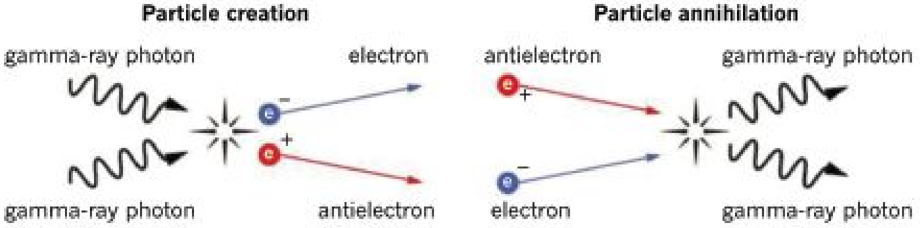


FIGURE 5.2.5 In the first 0.0001 seconds matter–antimatter pairs of particles were rapidly going in and out of existence. Here, two photons first produce an electron–positron pair but then an electron–positron pair annihilate, producing more photons.

As the temperature of the expanding universe fell to about 10^{12} degrees and became too low for the creation of new particles, there was therefore a great amount of annihilation of matter with antimatter and a huge reduction in the total amount of matter. The annihilation also produced an enormous increase in the amount of radiation. This huge fireball of radiation filled all space and dominated the universe for the next few hundred thousand years. It was, of course, the origin of the CMB radiation seen today.

PHYSICSFILE

Antimatter

Antimatter particles are similar to ordinary matter particles, but have the opposite sign. An 'antielectron' is called a positron and is the same as an electron except that it is positively charged. An antiproton has the same mass and other properties as a proton, but it has a negative charge. When a matter particle and an antimatter particle come together they annihilate each other, releasing all their mass as high-energy photons (light). Theoretically we could construct ordinary objects out of antimatter, but as soon as they came into contact with normal matter the two would annihilate with the release of absolutely huge amounts of energy.

This is discussed in more detail in Section 5.3.

Life's building blocks emerge

In the first few seconds while the temperature remained over a billion degrees or so, elementary particles known as **quarks** combined to form protons and neutrons. Protons and neutrons were forced close enough to fuse (join) together, forming hydrogen, helium and lithium nuclei. After a few more minutes the temperature dropped below that needed for this fusion and no further nuclei were formed (see Figure 5.2.6).

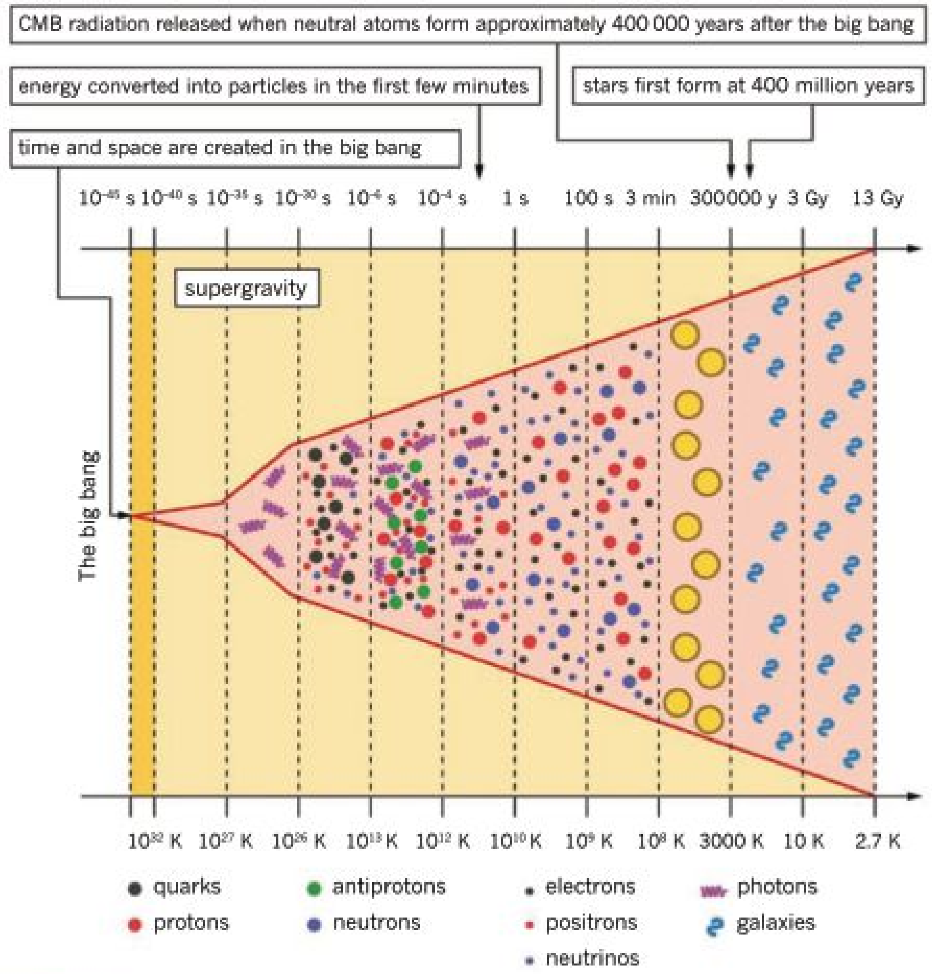


FIGURE 5.2.6 The 13.8 billion year history of the universe.

So how was any matter left after all the annihilation of matter and antimatter? Strangely, there was not quite an even balance of matter and antimatter in the early universe and so after the initial rapid annihilation, there was actually matter left over. The imbalance between matter and antimatter has been estimated at only about one extra matter particle in a billion particles—but that one in a billion now makes up our universe.

Without that very slight imbalance between matter and antimatter the universe would be empty—except for the CMB radiation.

For around 300 000 years the universe was too hot for atoms to form. Any electron captured by a proton was soon knocked away by an energetic photon. But by that time, the expansion of the universe resulted in the temperature dropping to about 3000 K. This meant photons were now in the red or infrared part of the spectrum. They did not have enough energy to ionise (strip electrons from) atoms—and so they continued filling space and became what we now call the CMB radiation. Now atoms of hydrogen, helium and lithium could form as these nuclei captured free electrons. Only these three nuclei formed in the first few minutes while the universe was hot enough for fusion to occur.

Formation of galaxies and stars

For the next billion years or so the universe was expanding, cooling and very dark. However, due to very slight irregularities in the original 'quantum foam' in which all those early particles came into existence, matter was starting to clump together. This process accelerated as the matter came closer and gravitational attraction increased.

The matter eventually started to form huge clumps, which are now the galaxies (refer back to Figure 5.2.6 on page 183 to put this stage into context). Within the galaxies matter collapsed into many smaller clumps. As these smaller clumps collapsed, energy was released when the matter (mostly hydrogen and helium) crashed together. This raised the temperature within the 'clumps' to millions of degrees. At these temperatures, the fusion (joining) of atoms was re-ignited—with hydrogen nuclei fusing to form new helium nuclei and releasing yet more heat. Stars were born.

These early stars were huge and very hot. They burnt out relatively quickly, and after using up all their hydrogen fuel they collapsed. This collapse, however, resulted in the release of yet more energy in huge supernovae explosions (see Figure 5.2.7). These explosions were so energetic that they resulted in the fusion of lighter elements into the heavier ones, a crucial process for the eventual formation of life.

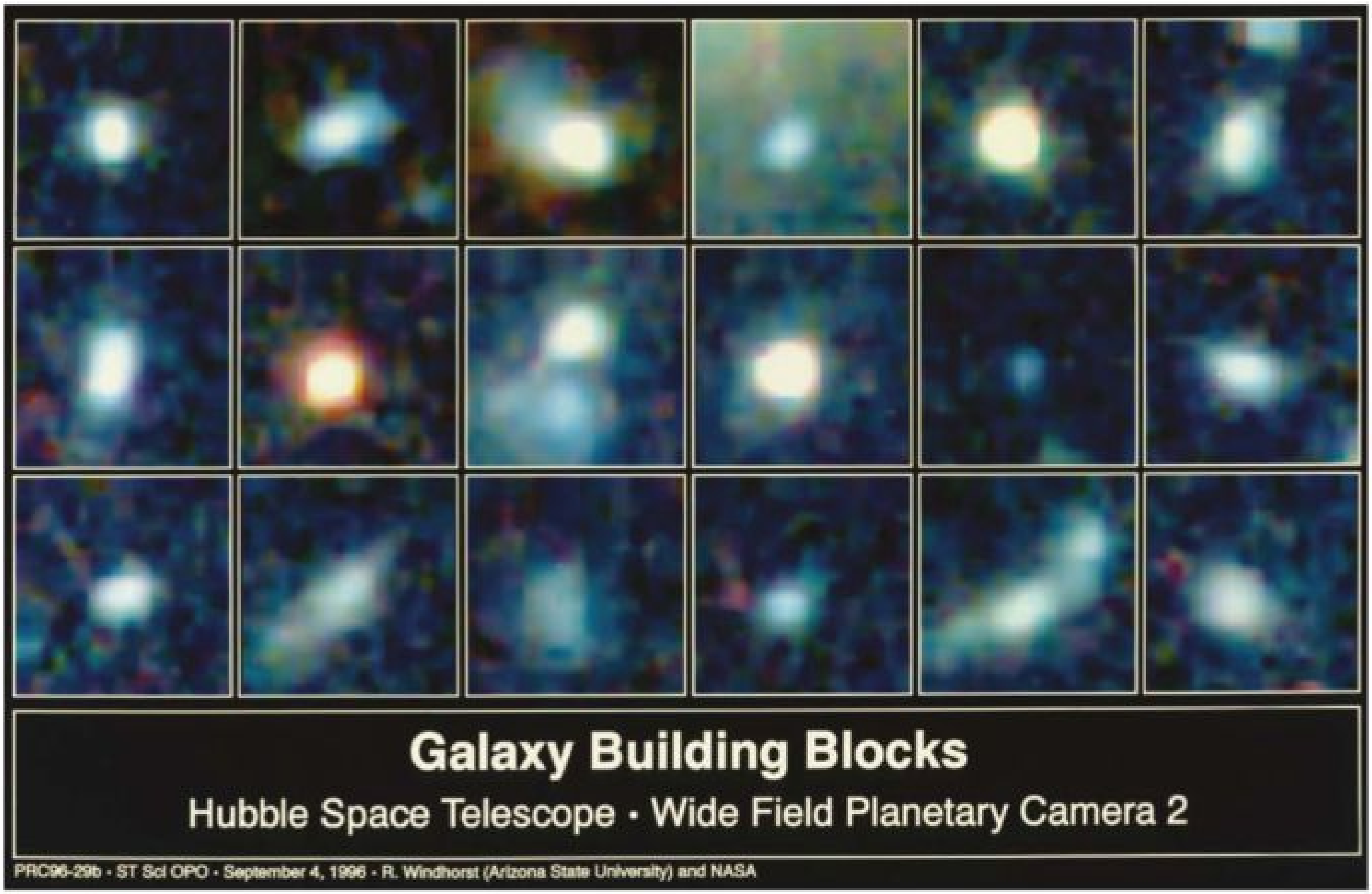


FIGURE 5.2.7 Early galaxies contained short-lived very hot (blue) stars. They collapsed as huge supernovae which resulted in the creation of the heavier elements.

Our Sun and solar system formed from the gravitational collapse of the 'dust' from some of these earlier supernova explosions and so contain heavier elements. Fortunately for us, the heavier elements were concentrated in the terrestrial (rocky) planets, including the Earth. And thus, eventually, life evolved.

5.2 Review

SUMMARY

- Last century, two different theories of the evolution of the universe seemed possible. Either the universe was in a 'steady state' with matter being continuously created or matter was all created with a 'big bang'.
- The big bang suggests that at one stage the enormous heat would have produced radiation which would now have cooled. The discovery of cosmic microwave background radiation strongly supported the big bang theory.
- Space, time and matter were all created at the big bang.
- Matter was created in an extremely short time called inflation. The expansion was so rapid that matter and antimatter pairs of particles could not annihilate until the expansion slowed.
- In the early universe there was slightly more matter than antimatter, and that is the matter that makes up the universe today.
- Galaxies and stars formed as the result of very small irregularities in the early universe. The irregularities allowed gravitational attraction to 'pull' matter together.
- The heavy elements necessary for life came later with the explosive supernova of heavy stars.

KEY QUESTIONS

1 Describe the cosmic microwave background radiation detected by Arno Penzias and Robert Wilson.

2 It was discovered that the universe appeared to be expanding. This would seem to imply it started small. How does the steady state theory explain the expansion?

3 Why is the discovery of cosmic microwave background radiation seen as supporting evidence for the big bang theory and not the steady state theory?

4 Describe the formation of matter in the early universe.

5 The inflation period in the early universe lasted only for an incredibly short time. What might have been a consequence had it lasted for a longer time?

6 What happens when an electron meets a positron?

7 When is cosmic microwave background radiation thought to have originated?

8 At only a few thousandths of a second after the big bang, matter and antimatter annihilated almost completely. What is the reason there was some matter left over?

9 All matter in the universe was created in the first few thousands of years and yet few of the elements essential for life existed then.

  a What elements did exist at that time and why?

  b How were the elements needed for life created?

5.3 Particles of the Standard Model

As a spin-off from the investigations into nuclear physics that were conducted towards the end of World War II, particle physicists began to predict and discover new subatomic particles. Sub-atomic particles are those smaller than an atom. Physicists recognised that as the energy of bombarding particles was increased, new particles were being formed.

At first, physicists built particle detectors and waited for high-energy cosmic rays from space to smash into their targets in order to see what nuclear fragments (pieces) they could identify. However, it was recognised that in order to probe more effectively into the nucleus, physicists needed to build more powerful machines such as particle accelerators. Particle accelerators accelerate protons and electrons to energies high enough to form the particles that the physicists were predicting would exist. Figure 5.3.1 shows the motion of particles detected in a particle accelerator.

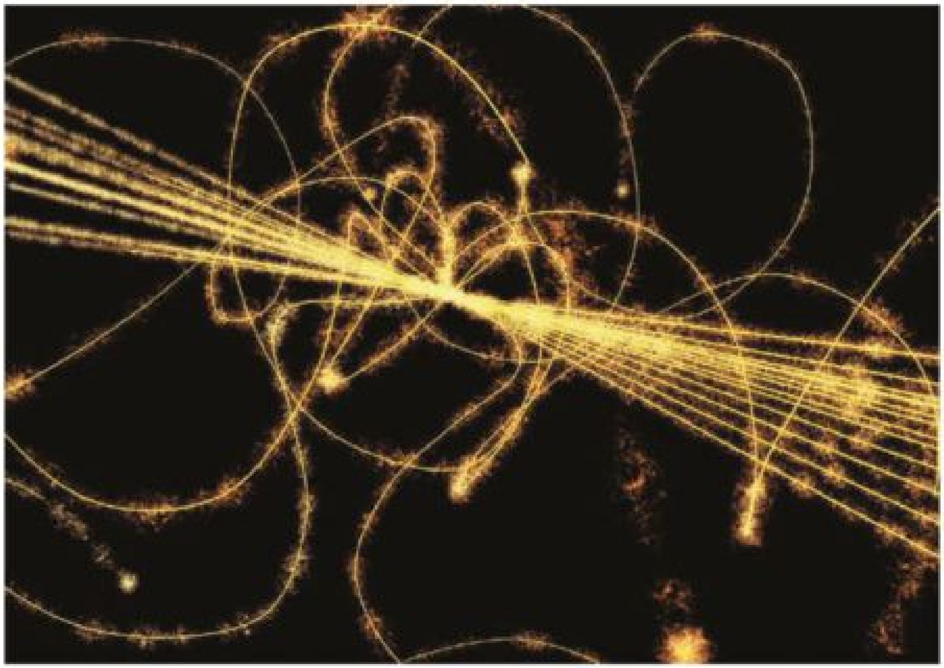


FIGURE 5.3.1 This image shows the paths of many particles in a particle accelerator. Energy is converted into mass as pairs of particles and their antiparticles are created from a gamma ray. The big bang theory and the Standard Model of particle physics tell us that this is how the matter inside us was created in the first seconds of the universe billions of years ago.

These new discoveries led to the development of the **Standard Model** of particle physics. The Standard Model is the best description so far of the fundamental building blocks of matter and the forces that govern them. This section will explore the particles of the Standard Model.

THE STANDARD MODEL OF PARTICLE PHYSICS

Currently, the Standard Model of particle physics is the most successful theory for predicting the behaviour and properties of the particles that exist in nature. The Standard Model was developed by experiment and theory together. As technology improved, new particles were found, and as the model developed, more particles were predicted and were then found. The Standard Model is a mathematical description of all known particles and three of the forces acting between them.

The essential components of the Standard Model are that forces between particles are mediated (caused) by other particles and that all matter is made of 12 fundamental particles.

The Standard Model consists of two types of particles: **fermions**, which make up all matter, and **gauge bosons**, which mediate the various interactions between other particles.

PHYSICS IN ACTION

Quarks and other subatomic particles

Our understanding of the atom has changed greatly in the past 100 years. It was once thought that atoms were like miniature billiard balls: solid and indivisible. The word 'atom' comes from the Greek '*atomos*', meaning indivisible. That idea was changed forever when the first subatomic particles—the electron, the proton and then the neutron—were discovered in the period from 1897 to 1932.

Since World War II, further research has uncovered about 300 other subatomic particles! Examples of these include pi-mesons, mu-mesons, kaons, tau leptons and neutrinos. For many years, physicists found it difficult to make sense of this array of subatomic particles.

Then in 1964 Murray Gell-Mann put forward a simple theory. He suggested that most subatomic particles were themselves composed of a number of more fundamental particles called quarks. Currently, it is accepted that there are six different quarks, each with different properties (and strange names!): up, down, charmed, strange, top and bottom. The latest quark to be identified was the top quark, whose existence was confirmed in 1995. The proton consists of two up quarks and one down quark, while neutrons consist of one up quark and two down quarks. Subatomic particles that consist of quarks are known as **hadrons**. The family of particles called the leptons has six members: electron, electron-neutrino, muon, muon-neutrino, tau and tau-neutrino. **Leptons** are indivisible point particles; they are not composed of quarks.

A significant amount of effort and money has been directed to testing Gell-Mann's theory—both theoretically and experimentally. This has involved the construction of larger and larger particle accelerators such as Fermilab in Chicago and CERN in Geneva.

While the current theory suggests that quarks and leptons are the ultimate fundamental particles that cannot be further divided, the nature of scientific theories and models is such that they can change as new experimental data are obtained. Are quarks and leptons made of smaller particles again? Time will tell!

FIGURE 5.3.2 The particle accelerator is at CERN, the European centre for high-energy physics. It accelerates protons from rest to 99.99995% of the speed of light in under 20 seconds.

GAUGE BOSONS

There are four forces that can act between particles and govern their behaviour:

- strong nuclear
- electromagnetic
- weak nuclear
- gravitational.

Each has a different strength and acts over different distances. For example, the **strong nuclear force** acts over very short distances (that is, on the subatomic scale), but is very large. Gravity acts over an infinite range, but is relatively weak.

The fundamental assumption in the Standard Model is that three of these forces (strong nuclear, electromagnetic and weak nuclear) arise through the exchange of particles called gauge bosons (or just bosons). Bosons are often called force-carrying, force-mediating or exchange particles.

Each force has its own boson. In this context, consider these forces as their particle equivalents, according to the notion of wave–particle duality. (Wave–particle duality is described in more detail in Unit 4 Physics.) In other words, consider these forces as particles.

PHYSICSFILE

The graviton

Theoretical physics predicts that there should be a boson for gravity too, called a graviton. However, this has not yet been found. Luckily the effects of gravity—although familiar to us on a human scale—are negligible at the subatomic scale, meaning the Standard Model works accurately without this missing piece.

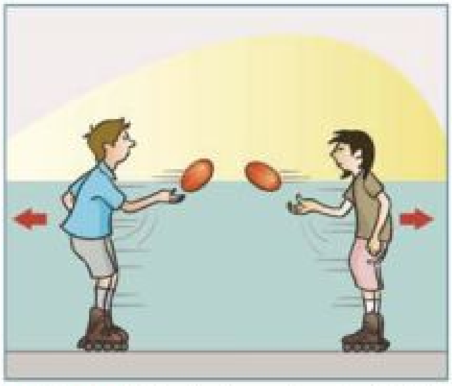


FIGURE 5.3.3 Inline skaters exchanging footballs act as an analogy for particles experiencing repulsive forces due to the exchange of gauge bosons.

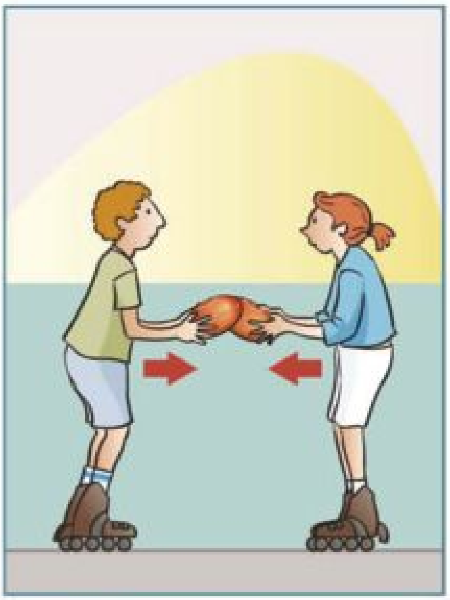


FIGURE 5.3.4 Inline skaters trying to grab footballs act as an analogy for particles experiencing attractive forces due to the exchange of gauge bosons.

Forces through exchange of particles

Previously, forces were thought of as being exerted on particles by fields. For example, there would be a region around a charged particle where another charged particle experiences a force. This may seem quite puzzling as the force is applied without any direct interaction by the two particles. The same effect is felt when two magnets are brought together. The magnets do not need to touch in order to feel the force between them. In the Standard Model this is resolved and forces are thought to be exerted through the exchange of other particles.

To use an analogy to see how a force can be exerted on two particles through the exchange of another particle, consider Figure 5.3.3. Two inline skaters stand stationary and then begin to pass footballs back and forth to each other. As they do this, they will begin to move away from each other. This is due to the conservation of momentum each time they throw and catch the ball. This situation could be likened to two particles experiencing a repulsive (pushing away) force.

A force of attraction can also be illustrated using the same analogy as shown in Figure 5.3.4. If the two skaters now exchange the footballs by trying to grab them out of each others' hands, they will exert a force of attraction on each other. This would cause them to move together and can be likened to two particles experiencing an attractive force.

Going back to a particle level, Figure 5.3.5 represents an electron approaching another electron. Each electron emits a photon that is absorbed by the other. This causes each electron to experience a force of repulsion. The photon has been responsible for the electromagnetic force acting on the two electrons.

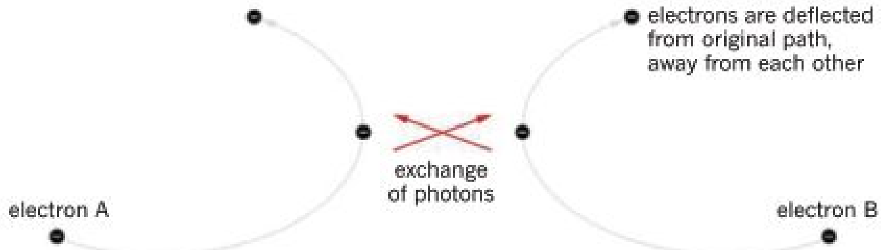


FIGURE 5.3.5 Two electrons approach each other, are repelled and then move away from each other. The two electrons exchange a photon that is the carrier of the electromagnetic force.

Table 5.3.1 gives a summary of the nature of these particles, their strength and the range over which they can exert a force.

Force	Nature	Relative strength	Range (m)	Force carrier (gauge boson)
strong nuclear	bonds nucleons together, acts between quarks	1	10^{-15} (~size of nucleus)	gluon
electromagnetic	responsible for both electric and magnetic fields exerting forces of attraction or repulsion	$\frac{1}{137}$	infinite	photon
weak nuclear	causes radioactive decay	10^{-6}	10^{-18} (less than the width of a proton)	W^+, W^- and Z
gravity	a force of attraction between any two objects with mass	6×10^{-39}	infinite	graviton (theoretical)

TABLE 5.3.1 Features of the various gauge bosons and the force they are responsible for.

FERMIONS

The Standard Model states that all particles of matter are comprised (made) of one or more of the 12 fundamental or elementary particles and their antimatter opposites. Being a fundamental particle means that, to the best of scientific knowledge, it is not comprised of other smaller particles. These fundamental particles are called fermions, which are defined by their **quantum numbers**.

Fermions are divided into two groups of particles called quarks and leptons. Six of the 12 particles are quarks and combine in groups of two or three to form the hundreds of known particles.

Quarks

All matter is made of particles comprised of protons, neutrons and electrons. However, protons and neutrons are actually made of even smaller particles called quarks. Quarks are the only things that interact using the strong nuclear force.

High-energy particle physicists seem to be a creative bunch: not only do they choose unusual names for their new particles but they also choose unusual ways to describe them and their properties. The six different types of quarks are referred to as the six 'flavours' of quarks.

The six flavours of quarks are:

- up (u)
- down (d)
- strange (s)
- charmed (c)
- bottom (b)
- top (t).

PHYSICSFILE

Naming quarks

The name 'quark' was taken by Murray Gell-Mann from the book *Finnegans Wake*, by James Joyce: 'three quarks for Muster Mark'.

FIGURE 5.3.6 This is an artist's representation of the three quarks that make up a proton (uud).

The last four quark names also apply to new quantum numbers and their conservation laws, called strangeness (*s*), charm (*c*), bottomness (*b*) and topness (*t*). Quarks have their antimatter opposites called antiquarks that have the opposite sign for all of their quantum numbers. Quarks have the properties of **baryon number**, spin and charge, which must add together to give the total baryon number, spin and charge of the particles that they combine to make (see Figure 5.3.6).

Table 5.3.2 shows the quantum numbers for quarks. (In this table, *e* is the fundamental or elementary charge introduced in Section 3.1. Its magnitude is equal to the charge on a proton.)

Particle/ antiparticle name	Symbol	Electro-magnetic charge (*Q*)	Baryon number (*B*)	Strangeness (*S*)	Charm (*c*)	Bottomness (*b*)	Topness (*t*)
up	u	$+\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	0
down	d	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	0	0
strange	s	$-\frac{1}{3}e$	$\frac{1}{3}$	−1	0	0	0
charmed	c	$+\frac{2}{3}e$	$\frac{1}{3}$	0	+1	0	0
bottom	b	e	$\frac{1}{3}$	0	0	−1	0
top	t	$+\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	+1
anti-up	$\overline{u}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	0	0	0
anti-down	$\overline{d}$	$+\frac{1}{3}e$	$-\frac{1}{3}$	0	0	0	0
anti-strange	$\overline{s}$	$+\frac{1}{3}e$	$-\frac{1}{3}$	+1	0	0	0
anti-charmed	$\overline{c}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	−1	0	0
anti-bottom	$\overline{b}$	$+\frac{1}{3}e$	$-\frac{1}{3}$	0	0	+1	0
anti-top	$\overline{t}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	0	0	−1

TABLE 5.3.2 The quantum numbers for quarks.

All quarks experience the strong nuclear force and this separates them from the leptons. Leptons do not experience the strong nuclear force. Quarks also have non-integer electromagnetic charges, meaning their electromagnetic charges are all fractions. Leptons have electromagnetic charges of −1 or 0.

When quarks bond together they form hadrons. Note that quarks have only been found bound into hadrons; they have never been found singly.

Hadrons

The common feature of the hadrons is that they all interact by exchanging **gluons**, which are the particles that mediate the strong nuclear force. Hadrons that carry an electromagnetic charge can also interact by exchanging photons, but the effect of the gluons far outweighs any other force-mediating particle.

This group of particles is further categorised into two groups, the **baryons** and the **mesons**, based on a property called baryon number (*B*), described below.

Mesons are made up of two quarks. Mesons are particles that have a baryon number of zero. In this group are many particles and their antimatter particles, for example the pion (π^+), antipion (π^-) and pi-zero (π^0), the kaon (K^+) and antikaon (K^-), and the eta (η^0). Note that the antimatter particle of the eta is considered to be itself.

Mesons consist of a quark and an antiquark pair. For example the pion-plus (π^+) consists of an up quark and an anti-down quark ($u\overline{d}$). Therefore the electromagnetic charge of the pion-plus is: $(+\frac{2}{3}) + (+\frac{1}{3}) = (+\frac{3}{3}) = +1$, but its baryon number is $(+\frac{1}{3}) + (-\frac{1}{3}) = (\frac{0}{3}) = 0$, as it should be for all particles that are not baryons.

Table 5.3.3 shows some of the many mesons and their constituent quarks.

Meson name	Symbol	Baryon number (*B*)	Strangeness (*S*)	Charm (*c*)	Bottomness (*b*)	Topness (*t*)	Quarks
pion-plus	π^+	0	0	0	0	0	$u\bar{d}$
pion-minus	π^-	0	0	0	0	0	$\bar{u}d$
kaon-plus	K^+	0	+1	0	0	0	$u\bar{s}$
kaon-minus	K^-	0	−1	0	0	0	$\bar{u}d$
rho-plus	ρ^+	0	+1	0	0	0	$u\bar{d}$
rho-minus	ρ^-	0	−1	0	0	0	$\bar{u}d$
phi	ϕ	0	0	0	0	0	$\bar{s}s$
D-plus	D^+	0	0	+1	0	0	$c\bar{d}$
D-zero	D^0	0	0	+1	0	0	$c\bar{u}$
D-plus-s	D_s^+	0	+1	+1	0	0	$c\bar{s}$
B-minus	B^-	0	0	0	−1	0	$b\bar{u}$
upsilon	Υ	0	0	0	0	0	$b\bar{b}$

TABLE 5.3.3 Various mesons and their constituent quarks.

Baryons are particles that have a baryon number of 1 for normal matter or −1 for antimatter. In this group are the familiar proton (p^+) and antiproton (p^-), neutron (n) and antineutron ($\bar{n}$), along with hundreds of other particles and their antiparticles, for example the lambda-zero (Λ^0) and antilambda-zero ($\overline{\Lambda}^0$), sigma-plus (Σ^+), sigma-zero (Σ^0) and sigma-minus (Σ^-), and the xi-zero (Ξ^0) and omega-minus (Ω^-).

Figure 5.3.7 shows the different groups of elementary particles covered so far and how the groups overlap.

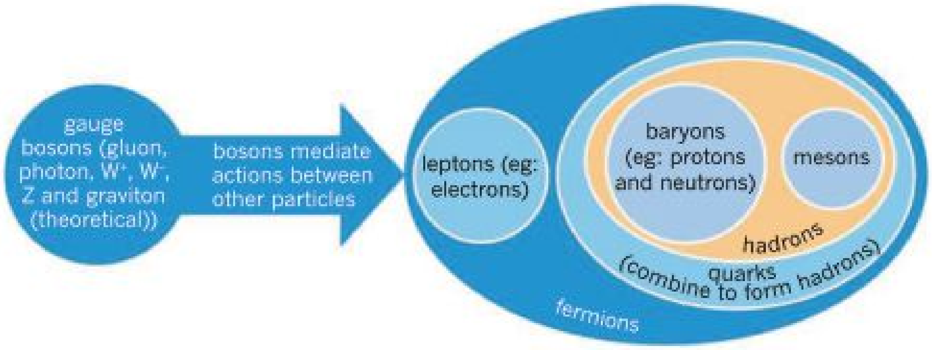


FIGURE 5.3.7 This diagram shows the different groups of elementary particles.

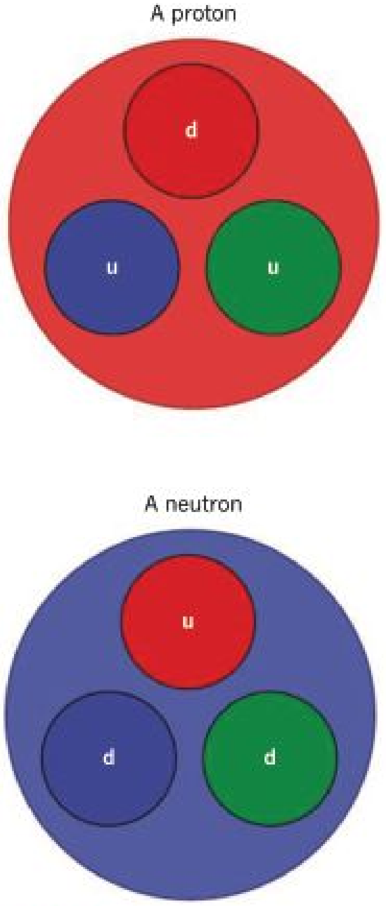


FIGURE 5.3.8 The proton and neutron are both baryons which are made of three quarks.

Baryons, including the proton and neutron, consist of three quarks. For example, the proton is made up of two up quarks and a down quark (uud) and the neutron is made up of one up quark and two down quarks (ddu), as shown in Figure 5.3.8.

Each of these quarks must have a different **colour charge** of red, green or blue. They must always bond together to give total colour charge of 'white', which means:

- red + blue + green (as with the proton and neutron shown above)
- antired + antiblue + antigreen
- red + antired
- blue + antiblue
- green + antigreen.

This is why there are never any 'free' individual quarks by themselves.

They must also combine to have the total electromagnetic charge of the proton, which is one 'fundamental unit' (e or $+1.60 \times 10^{-19}$ C). Therefore quarks must have an electromagnetic charge that is less that the fundamental unit. In fact quarks have an electromagnetic charge of either $+\frac{2}{3}e$ (up, charmed and top) or $-\frac{1}{3}e$ (down, strange and bottom), where e is the former fundamental unit of electromagnetic charge.

The proton's electromagnetic charge is therefore made up of:

$$(+\tfrac{2}{3}) + (+\tfrac{2}{3}) + (-\tfrac{1}{3}) = (+\tfrac{3}{3}) = +1$$

A neutron is made up of two down quarks and an up quark (ddu) of different colours, which equates to an electromagnetic charge of:

$$(-\tfrac{1}{3}) + (-\tfrac{1}{3}) + (+\tfrac{2}{3}) = (+\tfrac{0}{3}) = 0$$

Note that the baryon number of each quark is $+\frac{1}{3}$ so three quarks add to give a baryon number of +1. Antiquarks have a baryon number of $-\frac{1}{3}$. Adding the masses of the separate quarks in a proton is less than the mass of the proton. This is because most of the mass of the proton is stored in the gluons that hold the quarks together. Table 5.3.4 shows some of the many baryons and their constituent quarks.

Baryon name	Symbol	Baryon number (*B*)	Strangeness (*S*)	Charm (*c*)	Bottomness (*b*)	Topness (*t*)	Quarks
proton	p	+1	0	0	0	0	uud
antiproton	$\bar{p}$	−1	0	0	0	0	$\overline{uud}$
neutron	n	+1	0	0	0	0	udd
antineutron	$\bar{n}$	−1	0	0	0	0	$\overline{udd}$
Lambda-plus	Λ^+	+1	0	+1	0	0	udc
Lambda-zero	Λ^0	+1	−1	0	0	0	uds
Sigma-plus	Σ^+	+1	−1	0	0	0	uus
Sigma-zero	Σ^0	+1	−1	0	0	0	uds
Sigma-minus	Σ^-	+1	−1	0	0	0	dds
Xi-zero	Ξ^0	+1	−2	0	0	0	uss
Xi-plus	Ξ^+	+1	−2	0	0	0	dss
Omega-minus	Ω^-	+1	−3	0	0	0	sss

TABLE 5.3.4 Some of the many baryons and their constituent quarks.

MATTER AND ANTIMATTER

As mentioned with the particles discussed so far, most particles have what is called an antiparticle. These have the same properties of mass, spin, charge and lifespan as the particle. However, their electric charge and their quantum numbers are the same in magnitude but have the opposite sign. For example, a positron has the same mass as an electron, but has positive charge.

There are two conventions used to indicate an antiparticle.

- The antiparticle of uncharged particles are indicated by placing a bar above the symbol for the normal matter particle. An example is an electron neutrino. This particle has the symbol ν_e and its antiparticle, the anti-electron neutrino, has the symbol $\bar{\nu}_e$.
- The antiparticle of charged particles is given the symbol of the particle but with the opposite sign, e.g. the antiparticle of the muon (μ^-) is the anti-muon (μ^+).

Leptons

This particles in this group are the ones that interact by exchanging W and Z bosons, which mediate the **weak nuclear force**. Leptons that carry a charge can also interact by exchanging photons, which mediate the electromagnetic force. Leptons do not interact via the strong nuclear force carriers.

Electrons are a type of lepton. These particles also include the muon and tau particles, as well as their corresponding neutrinos and the antimatter opposites of these six particles. In your studies of nuclear physics, you will see that an electron, ejected from the nucleus in a beta-minus decay, will always be emitted along with an antineutrino. In a beta-plus decay, an antimatter positron is emitted with a normal neutrino.

The LEP (Large Electron–Positron collider) is the most powerful collider of leptons ever built. The ALEPH (Apparatus for LEP Physics) experiment determined the mass of the W and Z bosons and found that the number of particles with light neutrinos was three (see Figure 5.3.9).

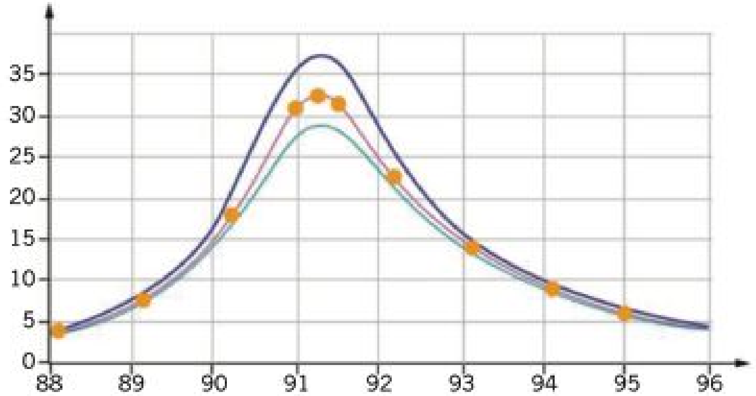


FIGURE 5.3.9 These three lines model the existence of two (green), three (red) or four (blue) neutrinos. The data from millions of Z particles, observed by the ALEPH experiment at CERN, produced the orange dots. There is a strong correlation between the data and the three-neutrino model.

When nuclei undergo beta-minus decay (β^-), an electron (e^-) is emitted from the nucleus, along with its corresponding electron antineutrino ($\bar{\nu}_e$). In a beta-plus decay (β^+), a positron (e^+) is emitted along with the electron neutrino (ν_e). **Neutrinos** are essentially massless particles that interact so weakly with matter that they can go through the entire Earth like photons (light) go through a pane of glass. Since neutrinos have zero charge they do not experience the electromagnetic force. Another lepton particle called the muon (μ^-) can also be emitted from the nucleus along with the muon antineutrino ($\bar{\nu}_\mu$) in a similar way to the beta particle. Muons are very similar to electrons, but 207 times larger. There is also an antimuon (μ^+) along with its associated muon neutrino (ν_μ). These muon neutrinos are found to be different to the electron neutrinos. The last leptons are the tau (τ^-) with its corresponding antitau neutrino ($\bar{\nu}_\tau$), and the antitau (τ^+) with its tau neutrino (ν_τ). The tau is extremely massive in comparison to the electron, at 3477 times its mass.

Table 5.3.5 gives a summary of leptons and their properties.

Lepton name	Symbol	Charge	Electron lepton number	Muon lepton number	Tau lepton number
electron	e^-	−1	−1	0	0
electron neutrino	ν_e	0	+1	0	0
muon	μ^-	−1	0	+1	0
muon neutrino	ν_μ	0	0	+1	0
tau	τ^-	−1	0	0	+1
tau neutrino	ν_τ	0	0	0	+1

TABLE 5.3.5 The properties of leptons.

Table 5.3.6 shows the six antileptons with their symbols, charges and conservation numbers.

Antiparticle name	Symbol	Charge	Electron lepton number	Muon lepton number	Tau lepton number
positron	e^+	+1	−1	0	0
electron antineutrino	$\bar{\nu}_e$	0	−1	0	0
antimuon	μ^+	+1	0	−1	0
muon antineutrino	$\bar{\nu}_\mu$	0	0	−1	0
antitau	τ^+	+1	0	0	−1
tau antineutrino	$\bar{\nu}_\tau$	0	0	0	−1

TABLE 5.3.6 The properties of antileptons.

Matter vs antimatter: Annihilation

When physicist Paul Dirac first proposed antimatter, he did more than just predict its existence. He also predicted that matter and antimatter will annihilate when they collide. This is observed and exploited on a daily basis in experiments and applications involving antimatter.

Figure 5.3.10 shows an antiproton entering along the track marked L (top), before colliding with a proton. When the proton and antiproton mutually annihilate, the huge amount of energy released produces several new particles whose tracks form the 'star' pattern seen in this image.

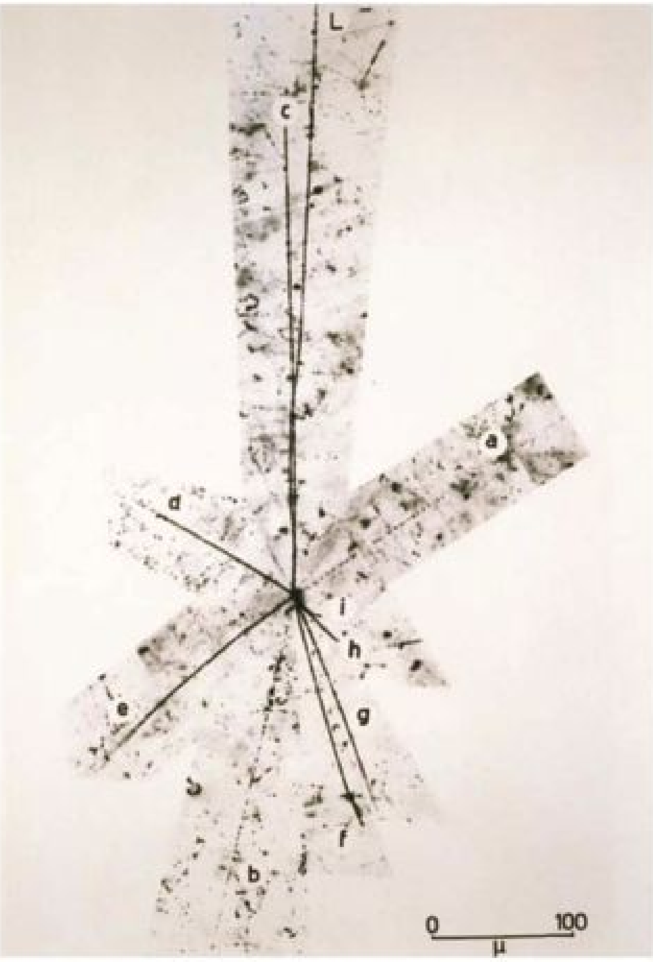


FIGURE 5.3.10 This image records a proton and antiproton annihilation. This event was recorded in 1955 in a photographic emulsion at the Bevatron accelerator at the Lawrence Berkeley Laboratory, California.

Another example is the annihilation of an electron and positron. This produces two photons that carry away the energy initially contained in the electron and positron. These events are a stunning verification of Einstein's famous equation $E = mc^2$, where mass (m) and energy (E) are equivalent and can be converted from one form to the other.

The opposite of annihilation is also observed. This is called particle–antiparticle pair production. Energy in the form of a photon can create a particle–antiparticle pair if the photon has energy greater than or equal to the mass of the particle–antiparticle pair. Pair production also illustrates the relationship between mass and energy in Einstein's equation.

PREDICTING THE EXISTENCE OF NEW PARTICLES

One of the most exciting aspects of physics is the discovery of something new. It could be a new theory or a more elegant solution to a problem. For particle physicists the ultimate achievement is to discover a new particle. These discoveries could unfold in one of two ways:

- the particle is predicted to exist through logical reasoning by theoretical physicists
- the particle is observed in a collision event by experimental physicists.

Whether by theory or observation, finding new particles generates great excitement and activity in the world of high-energy particle physics. Theoretical and experimental methods have been used over the years to discover, for example, the neutron, the neutrino, the positron and the Higgs boson.

The discovery of the neutron

The principle of conservation of momentum was used to interpret the data from experiments that led to the discovery of the neutron (see Figure 5.3.11). Because the neutron has no charge, it could not be investigated through the interactions of charged particles that had led to the discovery of the proton and electron.

In 1932, James Chadwick investigated collisions between alpha particles and the element beryllium. During the experiment Chadwick noticed that the calculations of conservation of momentum just didn't add up. Chadwick knew that the law of conservation of momentum was true, so he reasoned that there was an unknown particle involved. He figured that the particle had a mass close to the proton's mass, but without electric charge. Subsequent investigations confirmed his experiments and led to the naming of this particle as the neutron.

FIGURE 5.3.11 The first photographic evidence for the existence of the neutron, taken by French physicists Irene Curie and Frederic Joliot.

The discovery of the neutrino

Physicists studying the energy of the beta particles ejected from radioisotopes discovered that they seemed to vary by a small amount. They were missing a quantity of energy, which contravened (contradicted) the conservation laws of energy, spin and momentum. Wolfgang Pauli recognised that as the laws must not be violated there must be another very small particle produced in beta decay that carried the missing momentum and energy. The name given to this particle was the neutrino.

Neutrinos are particles that have been found to possess the lowest mass in nature, and they exist all throughout the universe. Neutrinos have no charge and their mass has only recently been discovered to be about one-billionth that of a proton (about 10^{-36} kg). While you have been reading these sentences, billions of neutrinos have passed right through your body, and continued on to pass right through the Earth. Fortunately neutrinos interact with matter very rarely and so are completely harmless. It has been estimated that if neutrinos passed through a piece of lead 8 light-years thick, they would still have only a 50% chance of being absorbed.

The discovery of the positron

Following the discovery of cosmic rays in the 1930s, a young physicist called Carl Anderson built an improved version of the cloud chamber that was being used by researchers to study the tracks of ionising radiation like cosmic rays. By placing his cloud chamber in a magnetic field, Anderson was able to photograph the curved path of charged ionising particles (like protons and electrons) that were created by bombarding cosmic rays.

Perhaps due to the improvements he had made to the design and the composition of the vapour, Anderson was able to collect very clear photographic evidence (see Figure 5.3.12). In one of his photos, Anderson noticed that one particle, with the same mass as an electron, curved in the opposite direction to what was expected due to the direction of the magnetic field it was passing through.

After further refinements of his experiment to exclude the possibility that the particle was a proton, Anderson and his colleagues concluded that it was indeed a positive equivalent of an electron. They dubbed it the positron. The antimatter positrons that Anderson discovered were created when a cosmic ray spontaneously created an electron–positron pair. For this discovery, Anderson was awarded the Nobel Prize in Physics at the age of 31.

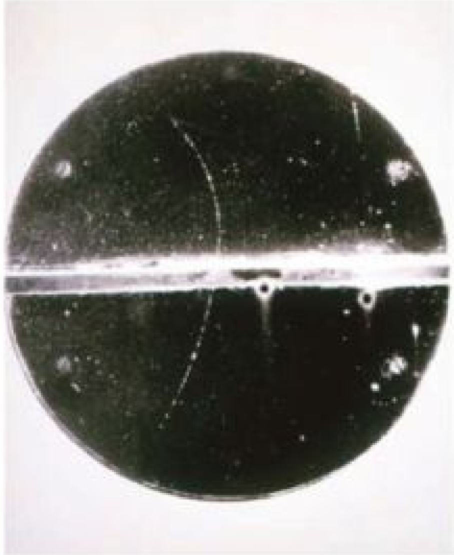


FIGURE 5.3.12 The discovery photograph of the positron by US physicist Carl Anderson. It shows the track of a positive particle that enters the chamber from below. The particle is known to be positive because of the way it bends in the chamber's magnetic field.

The discovery of the Higgs boson

Until the 1960s, physicists did not know how particles got their mass. British physicist Peter Higgs, and others, proposed an answer to this question in 1964. Using the Standard Model, they predicted that there was a yet-undiscovered particle that would interact with other particles and give them what is measured as mass. This particle came to be known as the Higgs boson.

The search for this, and for other particles, led over 10000 physicists and engineers to collaborate to build the **Large Hadron Collider** (LHC) near Geneva, Switzerland. This is the largest and most complex experimental facility ever built. One of its primary goals was to test the prediction of the existence of the Higgs boson.

The search continued until 2012, when evidence for the discovery of a candidate for the Higgs particle was first announced (see Figure 5.3.13). In 2013 two of the original researchers, Peter Higgs and François Englert, were awarded the Nobel Prize in Physics for their prediction and their work to prove the prediction.

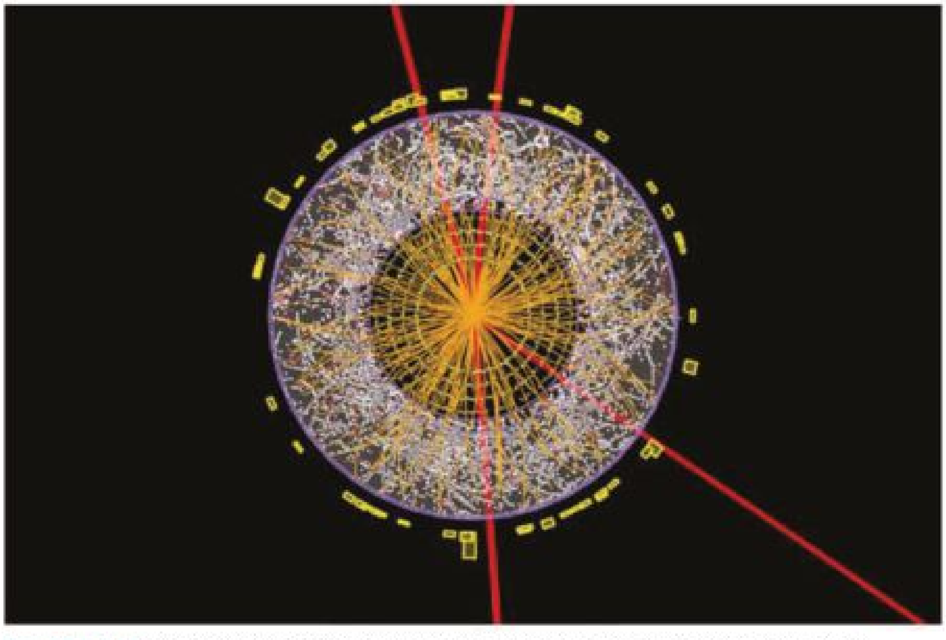


FIGURE 5.3.13 The image shows tracks of particles and measurements of their energies in the ATLAS detector at the LHC. The nature and energies of the particles produced are consistent with predictions of the formation of a Higgs boson.

5.3 Review

SUMMARY

- Our understanding of the universe progresses by both theory and experimentation. New particles are sometimes predicted before any experimental evidence is available, while sometimes observations in experimental events lead to the discovery of a new particle.
- The Standard Model of particle physics explains three of the four fundamental forces in the universe (electromagnetism, the strong nuclear force and the weak nuclear force) in terms of an exchange of particles called gauge bosons.
- The gauge bosons for these three forces are the photon, the gluon and the Z, W^- and W^+, for electromagnetism, the strong nuclear force and the weak nuclear force respectively.
- All matter in the universe is made of fundamental particles called quarks and leptons.
- There are six quarks which all experience the strong nuclear force mediated by gluons. These combine to form hadrons and cannot exist alone. The hadrons include baryons made of three quarks and mesons made of two quarks.
- Leptons are a group of particles that interact by exchanging W and Z bosons, which mediate the weak nuclear force. Electrons are in this group of particles.
- Antimatter particles have similar properties, like mass, spin and lifetime, as their corresponding particles. Their electric charge and other characteristics called quantum numbers are the same in magnitude but have an opposite sign.
- When a matter particle and its antimatter particle collide, they completely annihilate to produce a photon with the equivalent energy of the two particles, calculated by Einstein's mass–energy equation $E = mc^2$.

KEY QUESTIONS

1. Between which particles does the strong nuclear force act?
2. Which gauge bosons are responsible for the weak nuclear, strong nuclear and electromagnetic forces?
3. The particles of the Standard Model have been classified into three main groups. Name the three groups and give the main characteristic of each.
4. What is the main difference between the two groups of particles within the fermions (quarks and leptons)?
5. A common analogy used to explain how forces are mediated by the exchange of particles involves two skaters passing balls to each other. What does the ball represent in this analogy?
6. Name the two groups of hadrons and explain why they are separated.
7. Briefly explain two differences between leptons and quarks.
8. Classify each of the following as a gauge boson, a lepton or a hadron: proton, gluon, electron, photon, muon, neutron, neutrino.

Chapter review

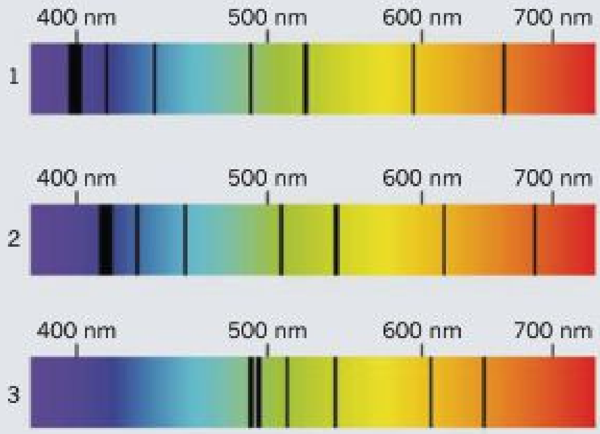
05

KEY TERMS

annihilation
antimatter
baryon
baryon number
big bang theory
colour charge
Doppler effect
fermion
gauge boson
gluon
hadron
Hubble Space Telescope
Hubble's constant
Hubble's law
intrinsic brightness
Large Hadron Collider
lepton
meson
nebula
neutrino
pair production
parallax movement
parsec
quantum mechanics
quantum number
quark
redshift
Standard Model
steady state theory
strong nuclear force
weak nuclear force

1 What causes redshift of light due to the Doppler effect?

2 Rank these spectra of galaxies in order of increasing distance of the galaxy from Earth and give your reasoning.



3 Hubble was able to work out the distance to many galaxies using the stars they contained called Cepheid variables. Why was this?

4 The modern value of the Hubble constant, H_0, is 67.80 km s^{-1} Mpc^{-1}. What is the correct interpretation of the meaning of the Hubble constant?

5 In what way is the expansion of the universe different from a conventional explosion?

6 A group of astronomers repeated Hubble's measurements to verify Hubble's law. The slope of the line of best fit they obtained for the graph of velocity of recession vs distance was steeper than Hubble's. What would they deduce about their values for the Hubble constant and for the age of the universe?

7 If the Hubble constant is 70 km s^{-1} Mpc^{-1}, what is the age of the universe in years to two significant figures?

8 For a galaxy 100 Mpc away, calculate how long it would take for light from that galaxy to reach us. The speed of light, $c = 3 \times 10^8$ m s^{-1}. Give your answer in years, to two significant figures.

9 Complete the following patagraph by choosing the correct term from the options in brackets.

In order for the [**fusion/fastening**] of lighter nuclei to form the more complex nuclei of the heavier elements, [**moderate/extreme**] temperatures and pressures were needed. Although these conditions [**did/did not**] occur in the early universe, they did occur [**billions/hundreds**] of years later as the early huge stars burnt out and collapsed due to [**electromagnetic/gravitational**] forces. This resulted in a supernova explosion which created those extreme temperatures and pressures and resulted in the production of [**heavier/lighter**] elements which were scattered into space by the explosion.

10 Describe the changes to CMB radiation over time.

11 What happens in the process called pair production ?

12 What happens in the process called annihilation?

13 Which quarks make up protons and neutrons?

14 Which forces are explained by the Standard Model of particle physics?

15 Explain how the proton gains a +1 charge from its constituent quarks and how a neutron has a neutral charge from its constituent quarks.

16 A neutron in an evacuated container decays to produce a proton and two other particles. The proton is then attracted to an electron and becomes the nucleus of a hydrogen atom. The atom then slowly drifts to the base of the container.

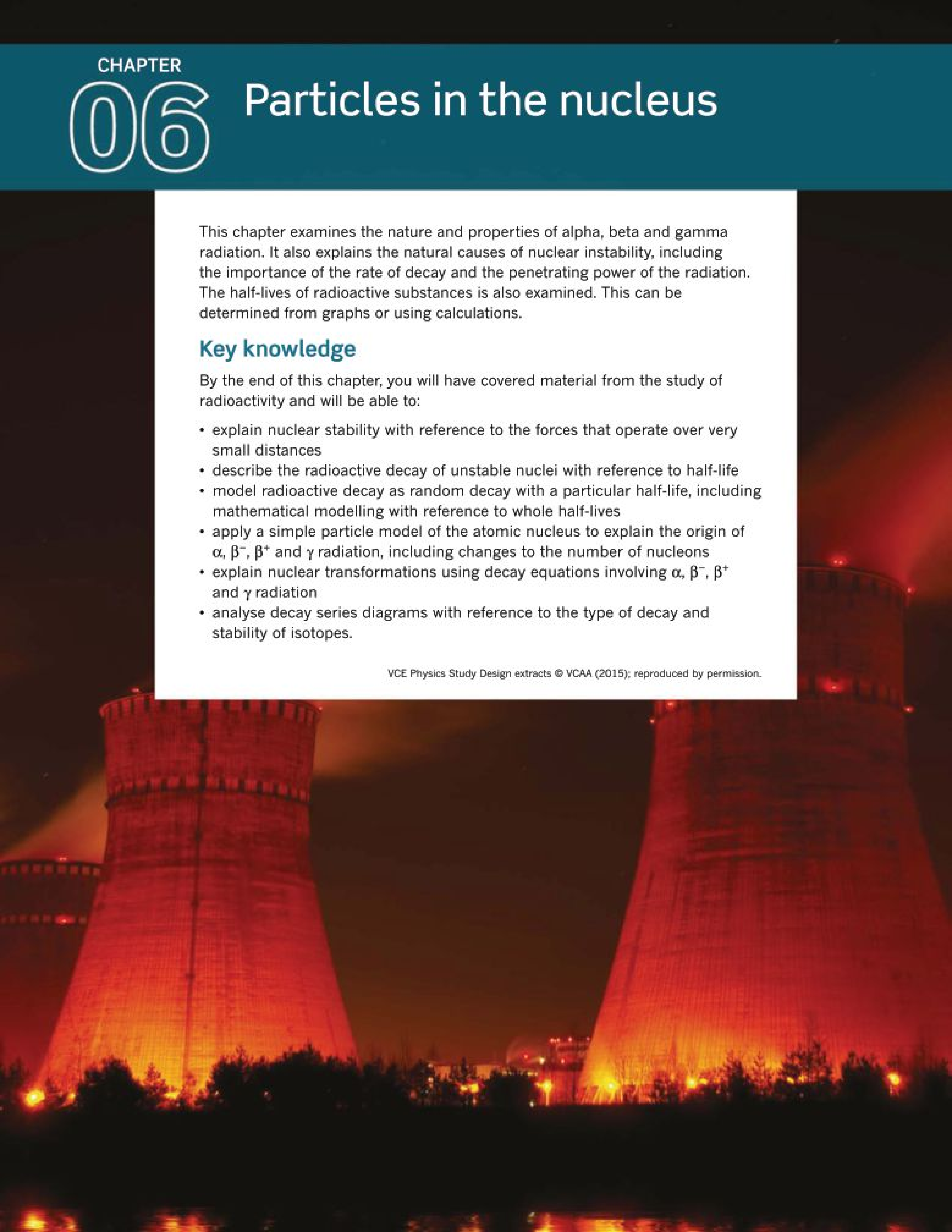
Order the four fundamental forces that have been involved in the sequence of events described above from first to last. If any were not involved then they should go last.

17 In your own words, explain how the Standard Model explains three of the four fundamental forces. As part of your answer, identify the three forces and the associated particle(s).

18 Is an electron a composite or a fundamental particle and is it a lepton or a hadron?

CHAPTER 06

Particles in the nucleus



This chapter examines the nature and properties of alpha, beta and gamma radiation. It also explains the natural causes of nuclear instability, including the importance of the rate of decay and the penetrating power of the radiation. The half-lives of radioactive substances is also examined. This can be determined from graphs or using calculations.

Key knowledge

By the end of this chapter, you will have covered material from the study of radioactivity and will be able to:

- explain nuclear stability with reference to the forces that operate over very small distances
- describe the radioactive decay of unstable nuclei with reference to half-life
- model radioactive decay as random decay with a particular half-life, including mathematical modelling with reference to whole half-lives
- apply a simple particle model of the atomic nucleus to explain the origin of α, β^-, β^+ and γ radiation, including changes to the number of nucleons
- explain nuclear transformations using decay equations involving α, β^-, β^+ and γ radiation
- analyse decay series diagrams with reference to the type of decay and stability of isotopes.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

6.1 Atoms, isotopes and radioisotopes

Many people mistakenly think that they never come into contact with radioactive materials or the radiation that these materials produce. However, the Earth is a radioactive planet and it is impossible to avoid exposure to radioactivity. Human senses cannot detect the radiation from radioactive atoms. High-energy radiation in higher than normal doses can be damaging to living tissue. Radiation and radioactive elements can also be used in a variety of applications that are beneficial. These radioactive atoms, or radioisotopes, will be discussed in this section.

ATOMS

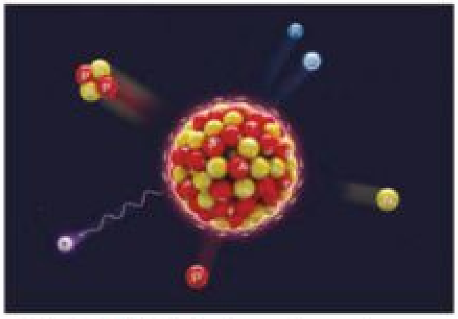


FIGURE 6.1.1 Radiation is spontaneously emitted from a radioactive nucleus.

If an atom is **radioactive**, it will spontaneously emit **radiation** from its nucleus. Figure 6.1.1 shows this radiation emitted in the form of particles and electromagnetic energy (light). To understand radiation and radioactivity, it is necessary to know about the structure of the atom. The central part of an atom, the **nucleus**, consists of particles known as **protons** and **neutrons**. Collectively, these particles are called **nucleons** and are almost identical in mass and size.

The nucleons have very different electrical properties. Protons have a positive charge. Neutrons are electrically **neutral** so they have no charge. The nucleus contains nearly all of the atom's mass.

Most of the atom is empty space occupied only by negatively charged particles called **electrons**. These are much smaller and lighter than protons or neutrons. Figure 6.1.2 shows the structure of a typical atom.

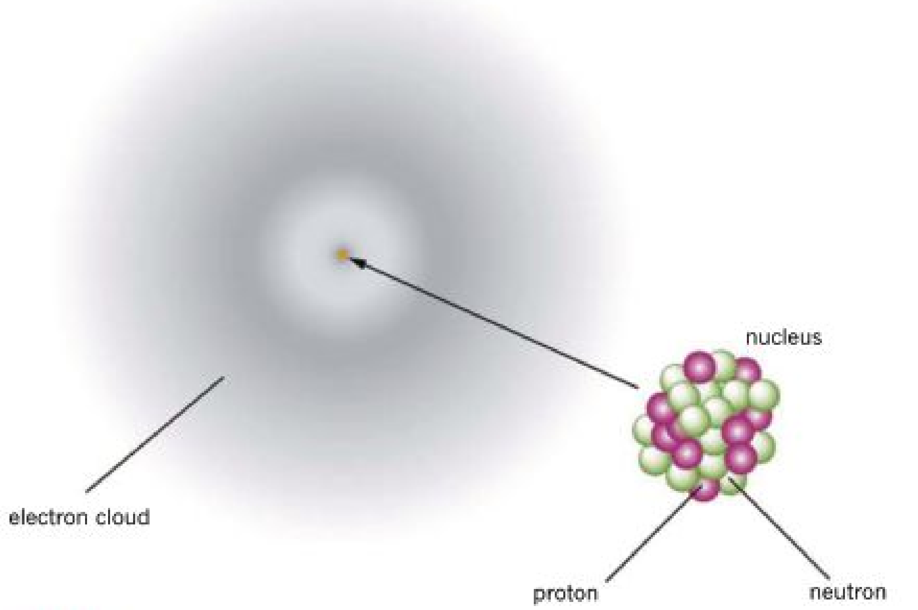


FIGURE 6.1.2 The nucleus of an atom occupies about 10^{-12} of the volume of the atom, yet it contains more than 99% of its mass. Atoms are mostly empty space. (Note, this atom is not drawn to scale.)

A particular atom can be identified by using atomic symbols that have the format shown in Figure 6.1.3.

FIGURE 6.1.3 Atomic notation.

The **mass number** (A) is the total number of protons and neutrons in the nucleus.

The **atomic number** (Z) is the number of protons in the nucleus.

Atoms with the same number of protons belong to the same element. For example, if an atom has six protons in its nucleus (i.e. $Z = 6$), then the atom must be carbon. The number of neutrons does not affect which element the atom is, but it does affect the mass of the atom. Figure 6.1.4 shows how the size of the nucleus depends on the mass number. The more protons and neutrons there are in a nucleus, the heavier and larger it is.

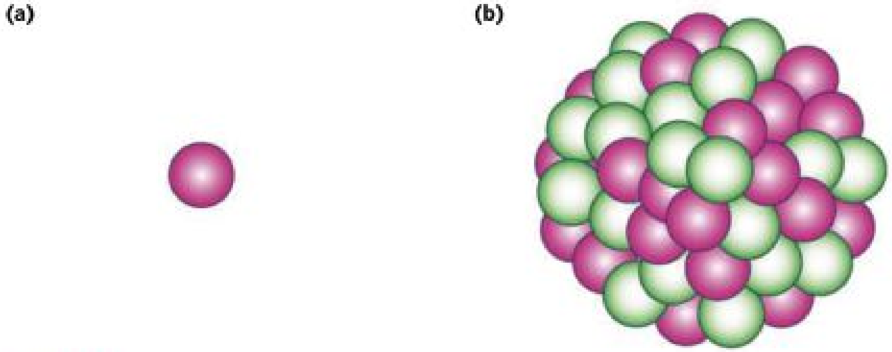


FIGURE 6.1.4 (a) and (b) are both nuclei, however the hydrogen atom (a) has a very different size nucleus to a uranium atom (b). (Note, the nuclei are not drawn to scale.)

In an electrically neutral atom the number of electrons is equal to the number of protons. For example, any neutral atom of uranium ($Z = 92$) has 92 protons in the nucleus and 92 electrons in the electron cloud.

PHYSICSFILE

Neutron stars

In the universe there are objects whose density is almost equal to that of nuclear matter. These are called neutron stars. They are gigantic balls with a radius of ten or more kilometres, and are made only of neutrons—something like a gigantic atomic nucleus. If filled up with this type of matter, a one-litre juice carton would weigh a thousand times more than the largest Egyptian pyramid.

ISOTOPES

All atoms of a particular element will have the same number of protons, but may have a different number of neutrons. For example, lithium exists naturally in two different forms. One form has three protons and three neutrons. The other has three protons and four neutrons. These different forms of lithium are called **isotopes** of lithium. These isotopes are illustrated in Figure 6.1.5.

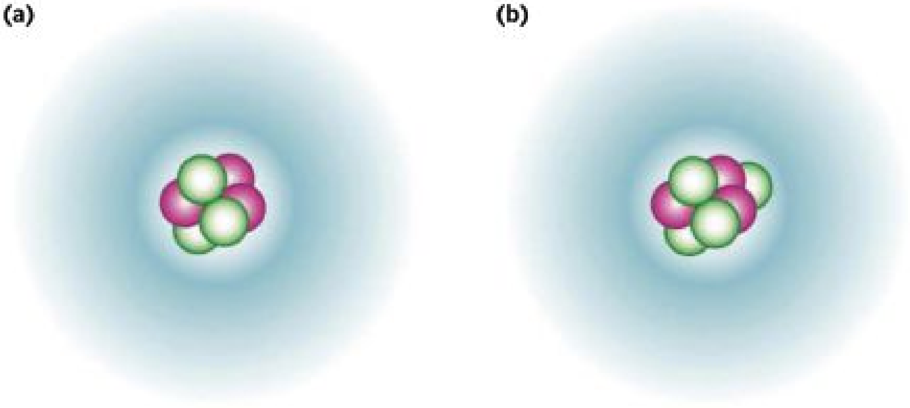


FIGURE 6.1.5 Two different isotopes of lithium: (a) $^{6}_{3}\text{Li}$ and (b) $^{7}_{3}\text{Li}$.

Isotopes are atoms that have the same number of protons but different numbers of neutrons. Isotopes have the same chemical properties but different physical properties such as density and volume.

The term **nuclide** is used when referring to a particular nucleus. For example, lithium-6 is a nuclide which has three protons and three neutrons.

There are three isotopes of hydrogen: the nuclide with one proton is called hydrogen, the nuclide with one proton and one neutron is called deuterium, and the nuclide with one proton and two neutrons is called tritium.

PHYSICSFILE

Heavy water

A compound of oxygen and deuterium has identical chemical properties to ordinary water. However, the molecular mass of ordinary water is about 18 (16 + 1 + 1) while the molecular mass of water containing deuterium is 20 (16 + 2 + 2). Thus, water that contains deuterium has a higher density (by about 11%) and is commonly known as 'heavy water'.

Worked example 6.1.1

WORKING WITH ISOTOPES

Consider the isotope of molybdenum, ${}^{95}_{42}\text{Mo}$. Work out the number of protons, nucleons and neutrons in this isotope.

Thinking	Working
The lower number is the atomic number, so this isotope has 42 protons.	atomic number = 42 This nuclide has 42 protons.
The upper number is the mass number. This indicates the number of particles in the nucleus, i.e. the number of nucleons.	mass number = 95 This nuclide has 95 nucleons.
Subtract the atomic number from the mass number to find the number of neutrons.	This isotope has 95 − 42 = 53 neutrons.

Worked example: Try yourself 6.1.1

WORKING WITH ISOTOPES

Consider the isotope of thorium, ${}^{230}_{90}\text{Th}$. Work out the number of protons, nucleons and neutrons in this isotope.

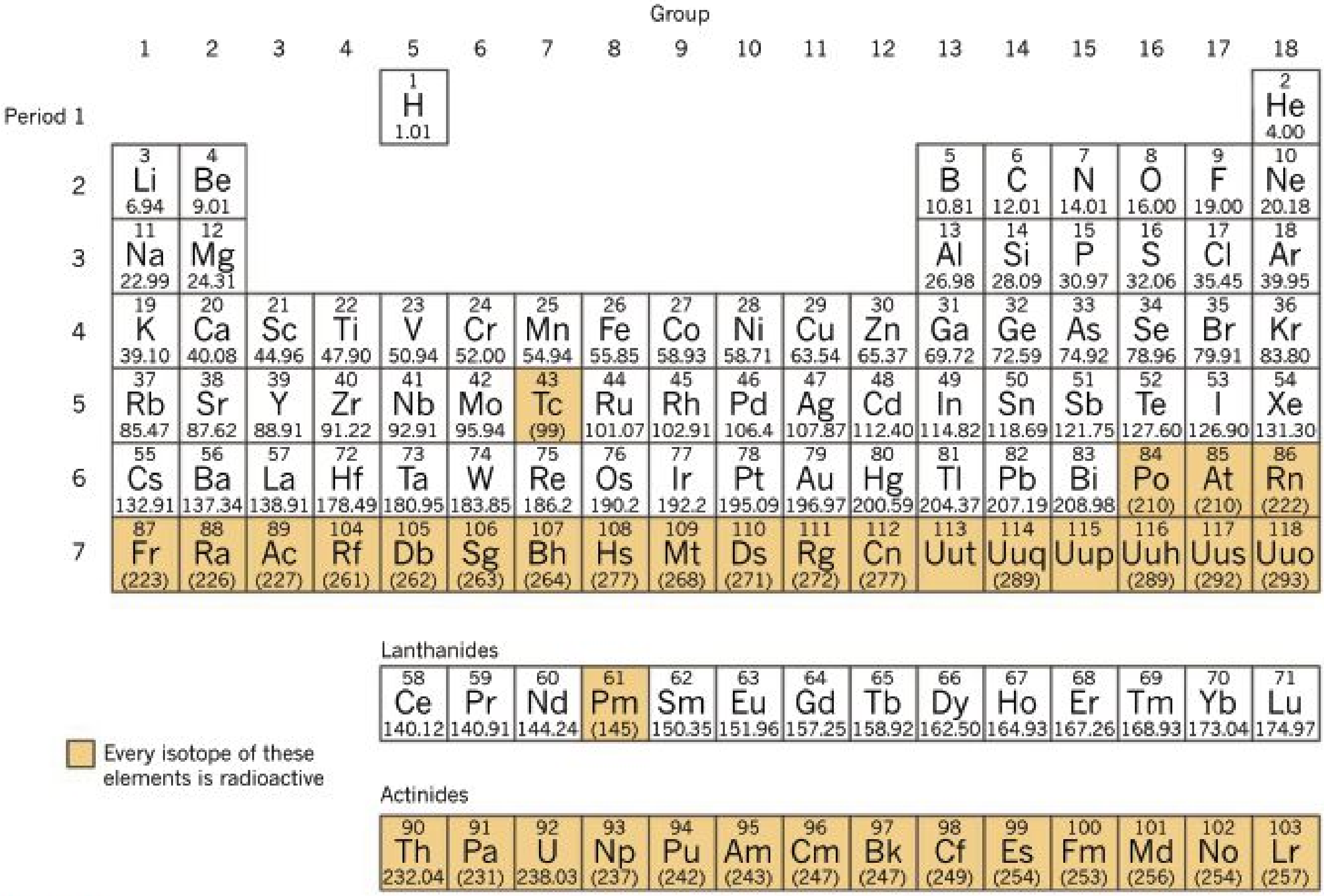
RADIOISOTOPES

Most of the atoms that make up the world around us are stable. Their nuclei have not altered in the billions of years since they were formed. These atoms will stay unchanged in the future. There are about 270 stable isotopes in nature. Tin ($Z = 50$) has ten stable isotopes, while aluminium ($Z = 13$) has just one.

There are also many naturally occurring isotopes that are unstable. An unstable nucleus may spontaneously become more stable by emitting a particle and so change into a different element or isotope. Unstable atoms are radioactive. An individual radioactive isotope is known as a **radioisotope**. Carbon has two stable isotopes: carbon-12 and carbon-13. Carbon also has one naturally occurring isotope that is unstable: carbon-14. The nucleus of a carbon-14 atom may spontaneously decay into a different substance, emitting high-energy particles that can be harmful. A known radioactive substance is identified by the radiation warning symbol shown in Figure 6.1.6.

FIGURE 6.1.6 This symbol is used to label and identify a radioactive source.

Figure 6.1.7 shows that every isotope of every element with an atomic mass greater than that of bismuth ($Z = 83$) is radioactive. The first 92 elements are naturally occurring.



Group

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Period 1					1 H 1.01													2 He 4.00
2	3 Li 6.94	4 Be 9.01											5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18
3	11 Na 22.99	12 Mg 24.31											13 Al 26.98	14 Si 28.09	15 P 30.97	16 S 32.06	17 Cl 35.45	18 Ar 39.95
4	19 K 39.10	20 Ca 40.08	21 Sc 44.96	22 Ti 47.90	23 V 50.94	24 Cr 52.00	25 Mn 54.94	26 Fe 55.85	27 Co 58.93	28 Ni 58.71	29 Cu 63.54	30 Zn 65.37	31 Ga 69.72	32 Ge 72.59	33 As 74.92	34 Se 78.96	35 Br 79.91	36 Kr 83.80
5	37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc (99)	44 Ru 101.07	45 Rh 102.91	46 Pd 106.4	47 Ag 107.87	48 Cd 112.40	49 In 114.82	50 Sn 118.69	51 Sb 121.75	52 Te 127.60	53 I 126.90	54 Xe 131.30
6	55 Cs 132.91	56 Ba 137.34	57 La 138.91	72 Hf 178.49	73 Ta 180.95	74 W 183.85	75 Re 186.2	76 Os 190.2	77 Ir 192.2	78 Pt 195.09	79 Au 196.97	80 Hg 200.59	81 Tl 204.37	82 Pb 207.19	83 Bi 208.98	84 Po (210)	85 At (210)	86 Rn (222)
7	87 Fr (223)	88 Ra (226)	89 Ac (227)	104 Rf (261)	105 Db (262)	106 Sg (263)	107 Bh (264)	108 Hs (277)	109 Mt (268)	110 Ds (271)	111 Rg (272)	112 Cn (277)	113 Uut	114 Uuq (289)	115 Uup	116 Uuh (289)	117 Uus (292)	118 Uuo (293)

Lanthanides

58	59	60	61	62	63	64	65	66	67	68	69	70	71
Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
140.12	140.91	144.24	(145)	150.35	151.96	157.25	158.92	162.50	164.93	167.26	168.93	173.04	174.97

Actinides

90	91	92	93	94	95	96	97	98	99	100	101	102	103
Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
232.04	(231)	238.03	(237)	(242)	(243)	(247)	(247)	(249)	(254)	(253)	(256)	(254)	(257)

🟧 Every isotope of these elements is radioactive

FIGURE 6.1.7 The periodic table of the elements.

PHYSICSFILE

Uuo

The element with the highest atomic number and highest atomic mass so far discovered is element 118, unumoctium (Uuo). Three atoms of this element were made in a particle accelerator in 2006 when calcium-48 nuclei were bombarded with californium-249 nuclei. The 20 protons of calcium combined with the 98 of californium to make just one or two atoms of Uuo. Uuo is very unstable and decays very rapidly, with a half-life of less than 1 ms.

Most of the elements found on Earth have naturally occurring radioisotopes; there are about 200 of these natural radioisotopes. During the twentieth century, an enormous number of radioisotopes were also artificially produced. Most of the radioisotopes used in industry, medicine and for scientific research are artificially produced. Artificial radioisotopes are produced in nuclear reactors or particle accelerators.

6.1 Review

SUMMARY

- The nucleus of an atom consists of positively charged protons and neutral neutrons. Collectively, protons and neutrons are known as nucleons. Negatively charged electrons surround the nucleus.
- The nucleus of the atom is extremely small but contains most of the atom's mass.
- The atomic number, *Z*, is the number of protons in the nucleus. The mass number, *A*, is the number of nucleons in the nucleus, i.e. the combined number of protons and neutrons. Elements are represented as $^{A}_{Z}X$.
- Isotopes of an element have the same number of protons but different numbers of neutrons. Isotopes of an element are chemically identical to each other, but have different physical properties.
- An unstable isotope—a radioisotope—may spontaneously decay by emitting a particle from the nucleus.

KEY QUESTIONS

1 What is the collective term used for protons and neutrons?

2 How many protons and how many neutrons are in the $^{197}_{79}Au$ nuclide?

3 How many nucleons are there in the $^{235}_{92}U$ nuclide?

4 Determine the number of protons, neutrons and nucleons in the following nuclides. You may need to refer to the periodic table in Figure 6.1.7 on page 203.

- **a** chlorine-35
- **b** plutonium-239

5 Which one or more of the following nuclides have seven neutrons in the nucleus? You may need to refer to the periodic table in Figure 6.1.7.

- **A** carbon-12
- **B** carbon-13
- **C** carbon-14
- **D** nitrogen-14

6 How is the number of electrons in a neutral atom determined?

7 Explain the meaning of the term isotope.

8 Krypton-84 is stable, but krypton-89 is radioactive. Imagine that you have just one atom of each isotope.

- **a** Are their atomic numbers and mass numbers the same or different? Justify your answers.
- **b** Compare the way these atoms would interact chemically with other atoms.

9 What is the difference between a stable isotope and a radioisotope?

10 Can a natural isotope be radioactive? If so, give an example of such an isotope.

6.2 Radioactivity

Around the turn of the twentieth century, scientists such as Marie Curie, pictured in Figure 6.2.1, were investigating the newly discovered radioactive substances polonium and radium. Ernest Rutherford and Paul Villard found that there were three different types of emission from these mysterious substances. They named them alpha, beta and gamma radiation.

Further experiments showed that the alpha and beta emissions were actually particles expelled from the nucleus. Gamma radiation was found to be high-energy electromagnetic radiation (light) also expelled from the nucleus. The term radioactive decay refers to the process that emits these particles and radiation from a nucleus.

The nature of these radiations will be discussed in this section.

FIGURE 6.2.1 Marie Curie, pioneer of research into radioactivity.

ALPHA (α) DECAY

When a heavy unstable nucleus undergoes radioactive decay, it may eject an **alpha particle**. This is a positively charged particle that consists of two protons and two neutrons. An alpha particle, symbol α, is identical to a helium nucleus and can also be written as ${}^{4}_{2}\text{He}$.

Uranium-238 is radioactive and may decay by emitting an alpha particle from its nucleus. Figure 6.2.2 shows the unstable nucleus of uranium-238 ejecting an alpha particle. This can be represented in a nuclear equation, which shows the changes occurring in the nuclei. Electrons are not considered in these equations, only the nucleons. The equation for the alpha decay of uranium-238 is:

$${}^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th} + {}^{4}_{2}\text{He} + \text{energy}$$

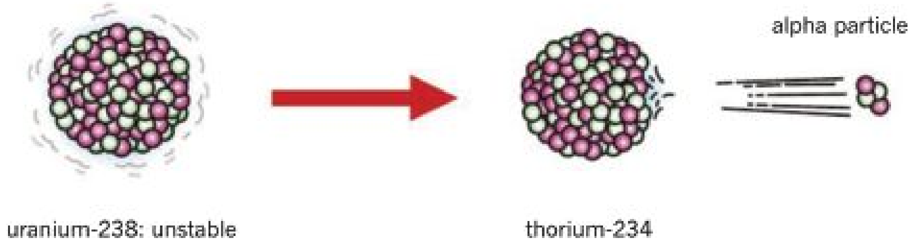


FIGURE 6.2.2 Alpha emission from uranium-238.

The **parent nucleus** ${}^{238}_{92}\text{U}$ has spontaneously emitted an alpha particle (α) and has changed into a completely different element, ${}^{234}_{90}\text{Th}$. Thorium-234 is called the **daughter nucleus**. The energy released is mostly kinetic energy carried by the fast-moving alpha particle.

When an atom changes into a different element, it is said to undergo a **nuclear transmutation**. In nuclear transmutations, electric charge is conserved. This results in the conservation of atomic number, i.e. the number of protons. The sums of atomic numbers on both sides of a nuclear equation must be equal. In the uranium decay equation, the atomic number is conserved: $92 = 90 + 2$. The mass number is also conserved: $238 = 234 + 4$.

In any nuclear reaction, including radioactive decay, atomic and mass numbers are conserved. Energy is released during these decays.

PHYSICSFILE

Radioactive lamps

The wicks or mantles used in old-style camping lamps are slightly radioactive. They contain a radioisotope of thorium, an alpha-particle emitter. They have not been banned from sale so far because they contain only small amounts of the radioisotope and can be used safely by taking simple precautions such as washing hands and avoiding inhalation or ingestion.

However, a scientist from the Australian National University in Canberra has called for these mantles to be banned because they tend to crumble and turn to dust as they age. If this dust were inhaled, alpha particles could settle in someone's lung tissue, possibly causing cancers to form.

FIGURE 6.2.3 An old radioactive gas-light mantle.

BETA (β) DECAY

Many radioactive materials emit **beta particles**. The there are two different types of beta particles: beta minus (β^-) and beta plus (β^+).

Beta minus (β^-)

This type of beta decay occurs when an electron is emitted from the *nucleus* of a radioactive atom, rather than from the electron cloud. This type of beta particle can be written as ${}^{0}_{-1}\beta$.

The atomic number of −1 indicates that the beta particle (the electron) has a single negative charge. The mass number of zero indicates that its mass is far less than that of a proton or a neutron.

Typically, beta-minus decay occurs if a nucleus has too many neutrons to be stable. A neutron spontaneously changes into a proton, a beta-minus particle (β^-, an electron), and an uncharged massless antimatter particle called an **antineutrino** ($\overline{v}$). This makes the nucleus more stable.

An example of an isotope that undergoes beta-minus decay is carbon-14. The other isotopes of carbon, i.e. carbon-12 and carbon-13, are both stable. Carbon-14 is unstable. It has too many neutrons and undergoes a beta-minus decay to become stable. In this process, one of the neutrons changes into a proton. Nitrogen-14 is then formed and energy is released. The beta-minus decay of carbon-14 is shown in Figure 6.2.4.

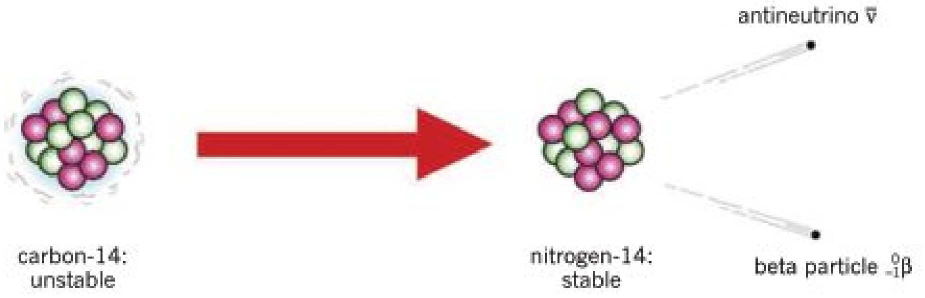


FIGURE 6.2.4 The beta-minus decay of carbon-14.

The nuclear equation for this decay is:

$${}^{14}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + {}^{0}_{-1}\beta + \overline{v} + \text{energy}$$

The transformation taking place inside the nucleus is:

$${}^{1}_{0}\text{n} \rightarrow {}^{1}_{1}\text{p} + {}^{0}_{-1}\beta + \overline{v}$$

Notice that, in all these equations, the atomic and mass numbers are conserved. The antineutrino has no charge and has so little mass that both its atomic and mass numbers are zero.

Beta plus (β^+)

A different form of beta decay occurs when a nucleus has too many protons. In this case, a proton may spontaneously change into a neutron and emit a neutrino (v) and a positively charged beta particle. This process is known as β^+ (beta-positive) decay. The positively charged beta particle is called a **positron**.

Positrons (${}^{0}_{+1}\beta$) have the same properties as electrons, but their electrical charge is positive rather than negative.

GAMMA (γ) DECAY

After a radioisotope has emitted an alpha or beta particle the daughter nucleus usually has excess energy. The protons and neutrons in the daughter nucleus then rearrange slightly and offload this excess energy by releasing a **gamma ray**, $^{0}_{0}\gamma$.

Gamma rays are high-energy electromagnetic radiation and so have no mass, are uncharged and travel at the speed of light (3.0×10^8 m s^{-1}).

A common example of a gamma ray emitter is iodine-131. It decays by beta and gamma emission to form xenon-131, as shown in Figure 6.2.5.

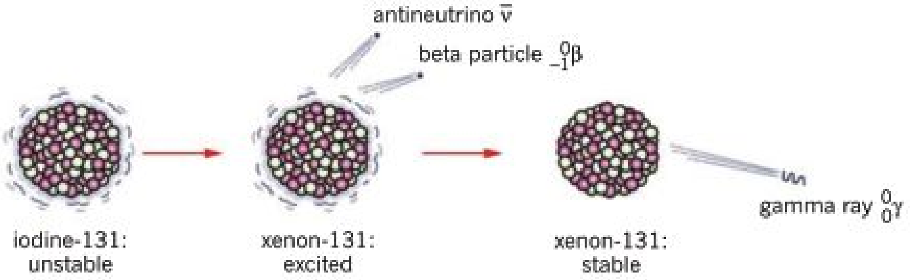


FIGURE 6.2.5 The gamma and beta decay of iodine-131.

The equations for this decay are:

$$^{131}_{53}\text{I} \rightarrow {}^{131}_{54}\text{Xe}^{*} + {}^{0}_{-1}\beta$$

$$^{131}_{54}\text{Xe}^{*} \rightarrow {}^{131}_{54}\text{Xe} + {}^{0}_{0}\gamma$$

The asterisk indicates that the nuclide is in an excited state.

Since gamma rays carry no charge and have no mass, they have no effect when balancing the atomic or mass numbers in a nuclear equation.

Worked example 6.2.1

RADIOACTIVE DECAY

Strontium-90 decays by radioactive emission to form yttrium-90. The equation is:
$^{90}_{38}\text{Sr} \rightarrow {}^{90}_{39}\text{Y} + \text{X}$

Determine the atomic and mass numbers for X and identify the type of radiation being emitted.	
Thinking	**Working**
The mass numbers of 90 are already balanced.	The mass number of X is zero.
Balance the atomic numbers.	$38 = 39 + a$ $a = 38 - 39 = -1$
X has atomic number of −1, and a mass number of zero.	X is a beta-minus particle, $^{0}_{-1}\beta$.

Worked example: Try yourself 6.2.1

RADIOACTIVE DECAY

Polonium-218 decays by emitting an alpha particle and a gamma ray.
The nuclear equation is:
$^{218}_{84}\text{Po} \rightarrow \text{X} + {}^{4}_{2}\alpha + \gamma$

Determine the atomic number and mass number for X, then use the periodic table to identify the element.

WHY RADIOACTIVE NUCLEI ARE UNSTABLE

In Figure 6.2.6 the stable nuclides that exist in nature are indicated by purple squares. The radioisotopes that are alpha and beta emitters can be identified by the black plus and minus symbols. Most of these radioisotopes also emit gamma radiation.

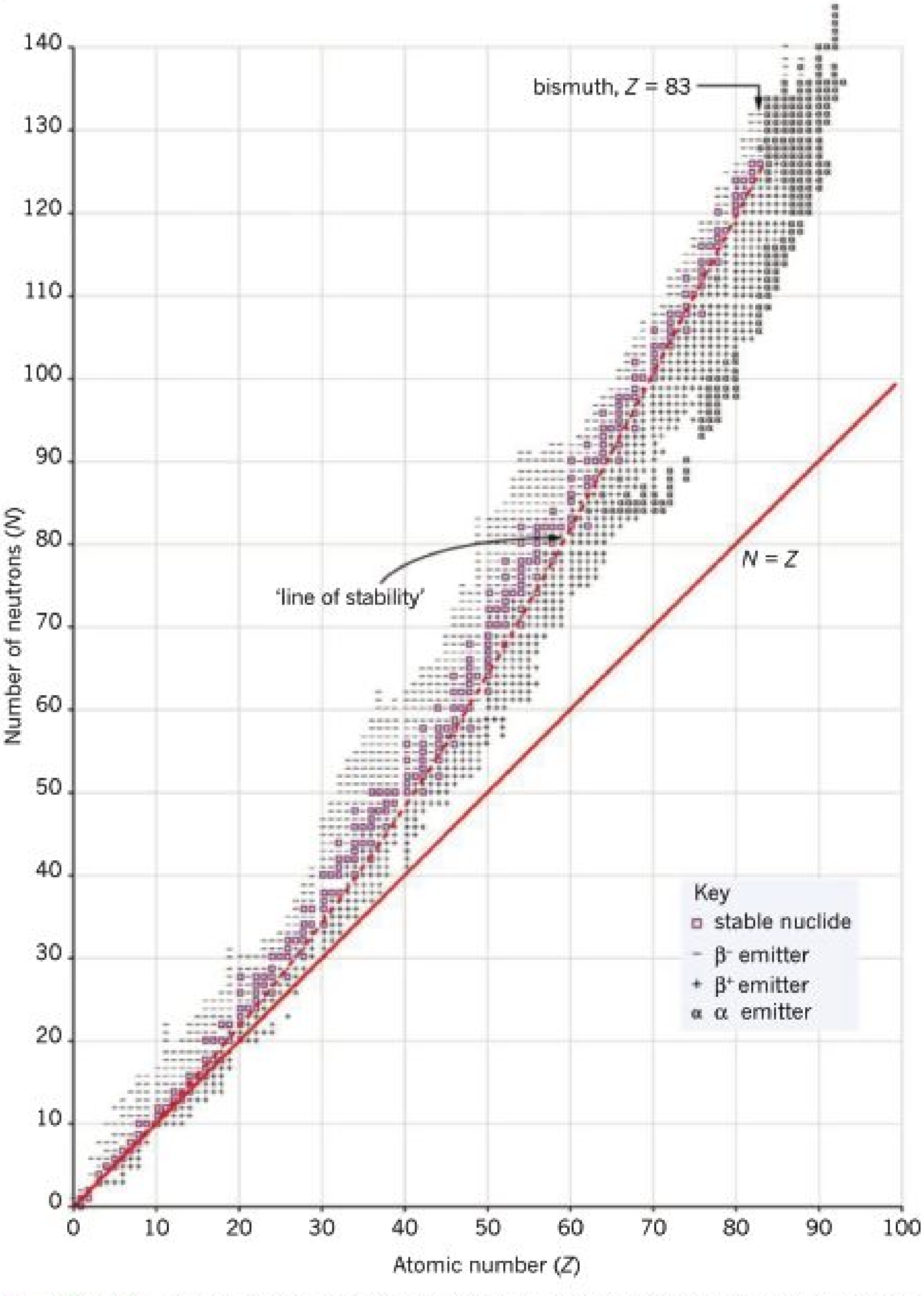


FIGURE 6.2.6 A chart showing stable and radioactive isotopes, plotted according to their number of protons (atomic number) and number of neutrons.

Within the nucleus, protons are in close proximity to other protons. This should seem odd since protons exert strong **electrostatic forces** of repulsion over each other. It might be thought that the nucleus should simply blow apart. As it doesn't, there must be other factors to consider. A force known as the strong nuclear force is also acting. This is a force of attraction that holds the nucleus together.

Electrostatic forces act between charged particles and can act over relatively large distances. In the nucleus, this means that each proton strongly repels every other proton so this force is trying to make the nucleus break apart. Neutrons are unaffected by electrostatic forces.

The strong nuclear force is a force of *attraction* that acts between every nucleon regardless of their charge. This force acts like 'nuclear glue'. Neutrons are attracted to nearby neutrons and protons. Protons are also attracted to nearby neutrons and protons. However, this force only acts over relatively short distances so for nucleons on the opposite sides of a large nucleus, this force is not significant.

In a stable nucleus, there is a delicate balance between the repulsive electric force and the attractive strong nuclear force. For example, bismuth-209, the heaviest stable isotope, has 83 protons and 126 neutrons. Here, the electrostatic repulsion of the protons is balanced by the strong attractive nuclear forces between the nucleons to make the nucleus stable. Compare this with bismuth-211. Its two extra neutrons upset the balance between forces. The nucleus of ^{211}Bi is unstable and it ejects an alpha particle in an attempt to become more stable.

From Figure 6.2.6 it is evident that there is a 'line of stability' (indicated by the curved red dashed line on the graph) along which the stable nuclei tend to cluster. Nuclei away from this line are unstable.

For small nuclei with atomic numbers up to about 20, the ratio of neutrons to protons in stable nuclei is close to one. However, as the nuclei become bigger, this ratio increases for stable nuclei. Zirconium ($Z = 40$) has a neutron-to-proton ratio of about 1.25, while for mercury ($Z = 80$) the ratio is close to 1.66. This indicates that for higher numbers of protons, nuclei must have even more neutrons to remain stable. These neutrons act to dilute the repelling forces that exist between the extra protons.

Elements with more protons than bismuth ($Z = 83$) simply have too many repulsive charges in the nucleus. Additional neutrons are unable to stabilise these nuclei. All of these elements are unstable and radioactive. Figure 6.2.7 illustrates stable and unstable nuclei.

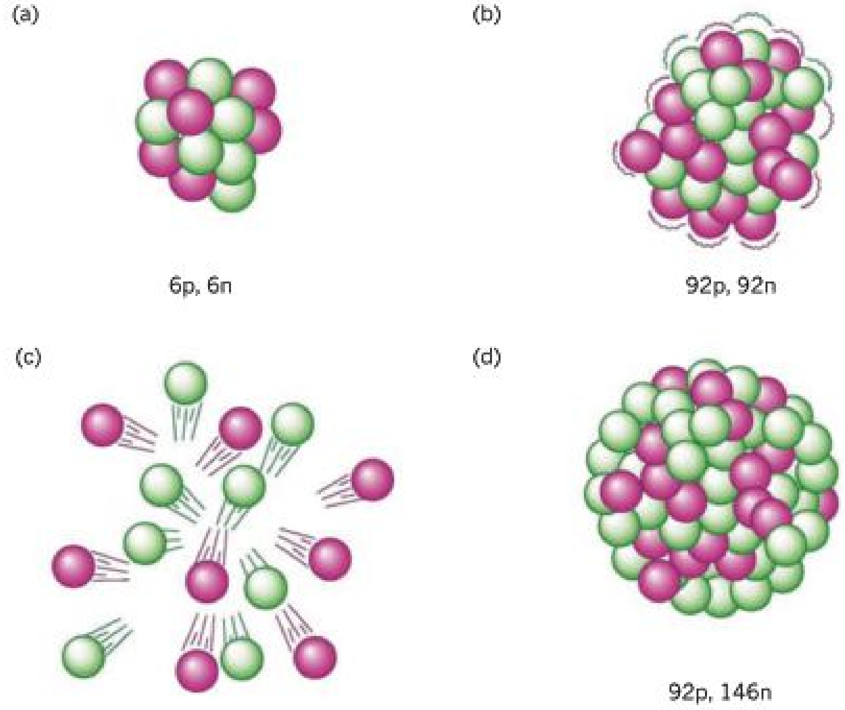


FIGURE 6.2.7 Stable and unstable nuclei. (a) A small nucleus such as carbon-12 is stable. This is because the electrostatic force of repulsion that acts between the protons is overcome by the strong nuclear force of attraction. (b) and (c) A large nucleus with equal numbers of protons and neutrons cannot exist. The electrostatic forces of repulsion between the protons would overcome the strong nuclear forces. (d) Additional neutrons increase the stability of large nuclei. The extra neutrons increase the influence of the strong nuclear force and act like a 'nuclear glue', holding the nucleus together.

EXTENSION

How radiation is detected

Our bodies cannot detect alpha, beta or gamma radiation. Therefore, a number of devices have been developed to detect and measure radiation.

A common detector is the **Geiger counter**. These are used:

- by geologists searching for radioactive minerals such as uranium
- to monitor radiation levels in mines
- to measure the level of radiation after a nuclear accident, such as the accident at Fukushima, Japan, in 2011
- to check the safety of nuclear reactors
- to monitor radiation levels in hospitals and factories.

A Geiger counter consists of a Geiger–Muller tube filled with argon gas, as shown in Figure 6.2.8.

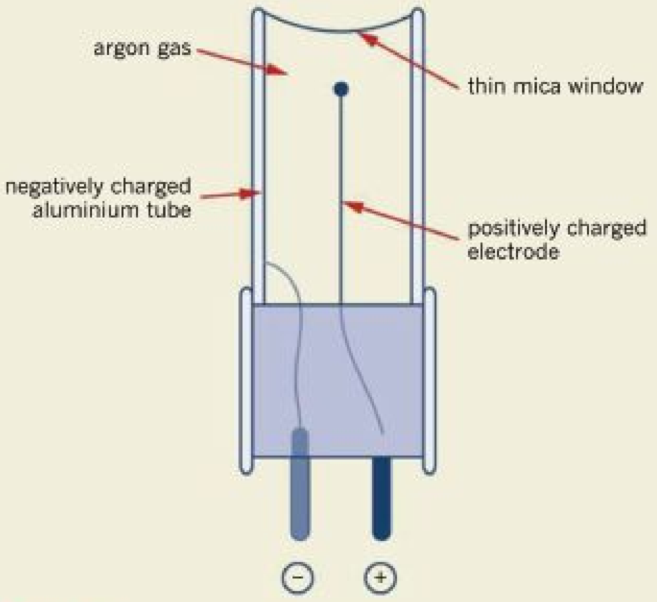


FIGURE 6.2.8 A schematic diagram of Geiger counter used for detecting ionising radiation.

A voltage of about 400 V is maintained between the positively charged central electrode and the negatively charged aluminium tube. When radiation enters the tube through the thin mica window, the argon gas becomes ionised and releases electrons. These electrons are attracted towards the central electrode and ionise more argon atoms along the way. For an instant, the gas between the electrodes becomes ionised enough to conduct a pulse of current between the electrodes. This pulse is registered as a count. The counter is often connected to a small loudspeaker so that the count is heard as a 'click' (see Figure 6.2.9).

FIGURE 6.2.9 A Russian scientist using a Geiger counter to measure radiation levels.

People who work in occupations that involve ongoing exposure to levels of ionising radiation usually pin a small radiation-monitoring device to their clothing. This is usually a thermoluminescent dosimeter (TLD), as pictured in Figure 6.2.10. These are used by personnel in nuclear power plants, radiotherapy departments at hospitals, airport security gates and uranium mines.

FIGURE 6.2.10 Thermoluminescent dosimeters are used by doctors, radiologists and scientists who work with radiation to monitor their exposure levels.

Thermoluminescent dosimeters contain a disc of lithium fluoride encased in plastic. Lithium fluoride can detect beta and gamma radiation, as well as X-rays and neutrons. TLDs are a cheap and reliable method for measuring radiation doses.

PHYSICS IN ACTION

How technetium is produced

Technetium-99m is the most widely used radioisotope in nuclear medicine. It is used for diagnosing and treating cancer. However, this radioisotope decays relatively quickly and so usually needs to be produced close to where it will be used. Technetium-99m is produced in small nuclear generators that are located in hospitals around the country (Figure 6.2.11). In this process, the radioisotope molybdenum-99, obtained from Lucas Heights, is used as the parent nuclide. Molybdenum-99 decays by beta emission to form a relatively stable (or metastable) isotope of technetium, technetium-99m, as shown below:

$$^{99}_{42}\text{Mo} \rightarrow {}^{99m}_{43}\text{Tc} + {}^{0}_{-1}\beta + \nu$$

Technetium-99m is flushed from the generator using a saline solution. The radioisotope is then diluted and attached to an appropriate chemical compound before being administered to the patient as a tracer. Technetium-99m is purely a gamma emitter. This makes it very useful as a diagnostic tool for locating and treating cancer. Its decay equation is:

$$^{99m}_{43}\text{Tc} \rightarrow {}^{99}_{43}\text{Tc} + \gamma$$

FIGURE 6.2.11 Technetium generators are used in hospitals that require radioisotopes. The generator has a thick lead shield that absorbs the beta and most of gamma radiation.

6.2 Review

SUMMARY

- Radioactive isotopes may decay by emitting alpha, beta or gamma radiation from their nuclei.
- An alpha particle (α) consists of 2 protons and 2 neutrons and is emitted from the nuclei of some radioisotopes. It is identical to a helium nucleus and can be written as $_{2}^{4}\mathrm{He}$.
- A beta-minus particle (β^- or $_{-1}^{0}\beta$) is an electron that has been emitted from the nucleus of a radioactive atom as a result of a neutron transmutating into a proton. An antineutrino is always emitted with a beta-minus particle.
- A beta-plus particle (β^+) or positron is a positively charged electron ($_{+1}^{0}\beta$) that has been emitted from the nucleus of a radioactive atom as a result of a proton transmutating into a neutron. A neutrino is always emitted with a beta-plus particle.
- A gamma ray, γ, is high-energy electromagnetic radiation that is emitted from the nuclei of radioactive atoms. Gamma rays always accompany an alpha or a beta emission.
- In any nuclear reaction, both the atomic and mass numbers are conserved.
- Nuclear stability is the result of the strong nuclear force between nucleons over a very short distance, which opposes electrostatic repulsion between protons.

KEY QUESTIONS

1 Determine the nature of the unknown, X, for the following transmutation:

$_{86}^{222}\mathrm{Rn} \rightarrow {}_{84}^{218}\mathrm{Po} + \mathrm{X}$

2 What type of beta decay occurs in this nuclear reaction?

$_{82}^{214}\mathrm{Pb} \rightarrow {}_{83}^{214}\mathrm{Bi} + \mathrm{Y}$

3 What type of decay occurs when a nucleus has too many protons?

4 What is a positron?

5 What is the nature (or identity) of alpha, beta and gamma radiation?

6 What are the mass numbers of the six stable nuclides of calcium (atomic number 20)? Use Figure 6.2.6 on page 208 to answer this question.

7 Where in an atom do alpha, beta and gamma radiation originate from?

8 For the unknown nuclides X and Y in each of these decay equations, determine the atomic number and mass number, and use the periodic table to identify the unknown elements.

a $_{92}^{238}\mathrm{U} \rightarrow \alpha + \mathrm{X} + \gamma$

b $_{88}^{228}\mathrm{Ra} \rightarrow \mathrm{Y} + \beta^- + \gamma$

9 Carbon-14 decays by beta-minus emission to form nitrogen-14. The equation for this is:

$_{6}^{14}\mathrm{C} \rightarrow {}_{7}^{14}\mathrm{N} + {}_{-1}^{0}\beta + \bar{\nu} + \text{energy}$

a How many protons and neutrons does the nitrogen atom have?

b What particle on the left side of the equation has transformed into what particle(s) on the right side of the equation?

10 What is missing in each of the following decay equations?

a $_{20}^{47}\mathrm{Ca} \rightarrow {}_{21}^{47}\mathrm{Sc} + ?$

b $_{70}^{150}\mathrm{Yb} \rightarrow {}_{68}^{146}\mathrm{Er} + ?$

6.3 Properties of alpha, beta and gamma radiation

In the early experiments with radioactivity, emissions from a sample of radium were directed through a magnetic field as shown in Figure 6.3.1. The emissions followed three distinct paths, which suggested that there were three different forms of radiation being emitted. The emissions each had different charges, masses and speeds. These properties will be discussed in this section.

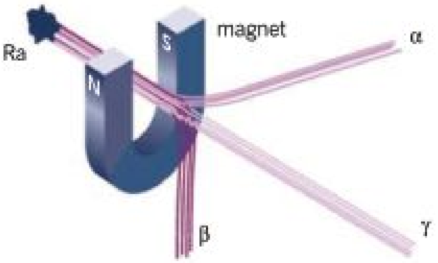


FIGURE 6.3.1 Applying a magnetic field shows that there are three different types of emissions from a radium source.

ALPHA (α) PARTICLES

Alpha particles consist of two protons and two neutrons. This means that they are relatively heavy and slow moving. Alpha particles are emitted from the nucleus at speeds of up to 20 000 km s^{-1} (2.0×10^7 m s^{-1}), just less than 10% of the speed of light (see Figure 6.3.2).

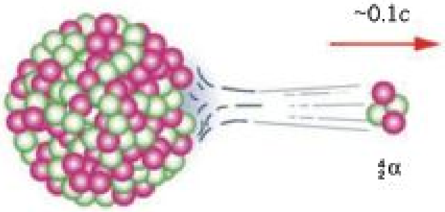


FIGURE 6.3.2 The speed and structure of an α particle.

Alpha particles have a double positive charge. This, combined with their relatively slow speed, makes them very easy to stop. They only travel a few centimetres in air before losing their energy, and will be completely absorbed by thin card or a human hand (see Figure 6.3.3). They have a poor **penetrating ability.**

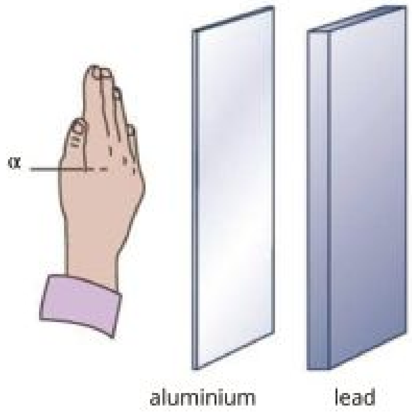


FIGURE 6.3.3 The penetrating ability of α radiation.

An example of an isotope that emits α radiation (or undergoes α decay) is the isotope of americium $^{241}_{95}\text{Am}$. Americium can be found in ionisation smoke detectors (see Figure 6.3.4).

$$^{241}_{95}\text{Am} \longrightarrow {}^{237}_{93}\text{Np} + {}^{4}_{2}\alpha$$

FIGURE 6.3.4 A domestic smoke detector contains a radioactive alpha emitter.

BETA (β) PARTICLES

Beta particles are fast-moving electrons (β^-) or positrons (β^+). Beta-minus particles are created when a neutron decays into a proton. Beta-plus particles are created when a proton decays into a neutron.

Beta particles are much lighter than alpha particles. As a result, they leave the nucleus with far higher speeds—up to 90% of the speed of light (*c*), as shown in Figure 6.3.5.

FIGURE 6.3.5 The speed and structure of a β particle.

Beta particles are more penetrating than α particles. They are faster and have a smaller charge than α particles. Beta particles will travel a few metres through air and through a human hand. Typically, a sheet of aluminium about 1 mm thick will stop them, as shown in Figure 6.3.6.

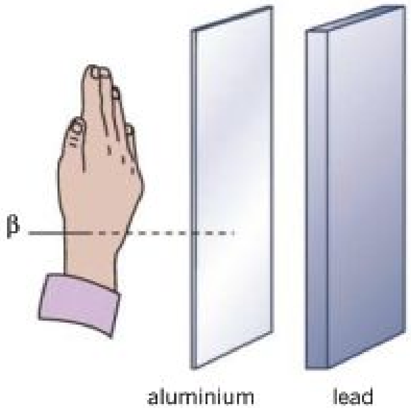


FIGURE 6.3.6 The penetrating ability of β radiation.

GAMMA (γ) RAYS

Gamma rays are electromagnetic radiation with a very high frequency. Figure 6.3.7 shows where gamma rays lie along the electromagnetic spectrum. They have no rest mass and travel at the speed of light: 3.0×10^8 m s^{-1} or 300 000 km s^{-1} (see Figure 6.3.8).

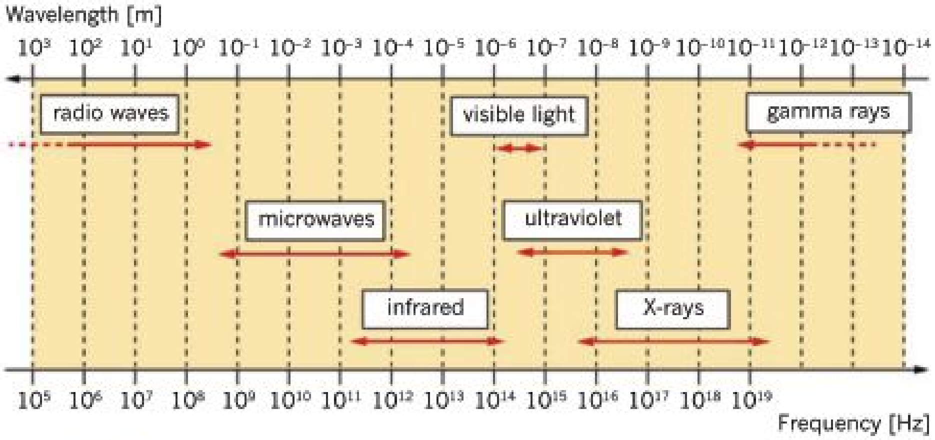


FIGURE 6.3.7 The electromagnetic spectrum contains many different types of radiation that differ in their wavelength and frequency. Gamma rays have very high frequencies and very short wavelengths, making them very energetic and highly penetrating.

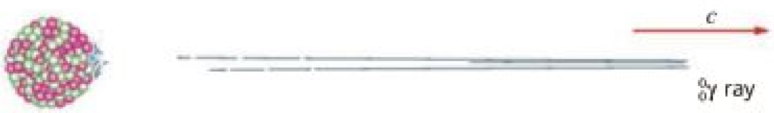


FIGURE 6.3.8 The speed and nature of γ radiation.

Gamma rays have no electric charge. Their high energy and uncharged nature make them a very penetrating form of radiation. Gamma rays can travel an almost unlimited distance through air and even through a human hand, an aluminium sheet and a few centimetres of lead (see Figure 6.3.9). Even a metre of concrete would not completely absorb a beam of γ-rays.

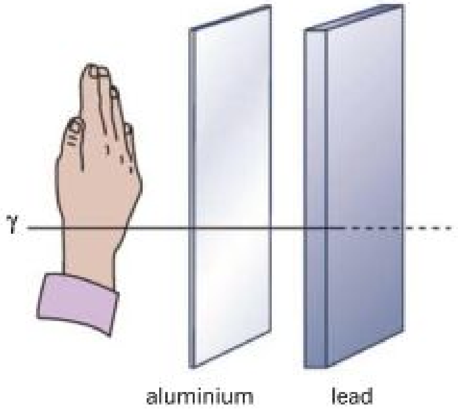


FIGURE 6.3.9 The penetrating ability of γ radiation.

PHYSICS IN ACTION

Monitoring the thickness of sheet metal

Beta particles can be used to monitor the thickness of rolled sheets of metal and plastic during manufacture, as shown in Figure 6.3.11. A β particle source is placed under the newly rolled sheet and a detector is placed on the other side. If the sheet being made is too thick, fewer β particles will penetrate and the detector count will fall. This information is instantaneously fed back to the rollers and the pressure is increased until the correct reading is achieved and hence the right thickness of metal is attained.

Alpha particles or gamma rays would not be appropriate for this task.

Alpha particles have a very poor penetrating ability, so none would pass through the metal. Gamma rays usually have a high penetrating ability and so a thin metal sheet would not stop them. In addition, workers would need to be shielded from γ radiation.

The penetrating properties of beta rays make them ideal for this job. The thickness of photographic film and plastic sheets is also monitored in this way.

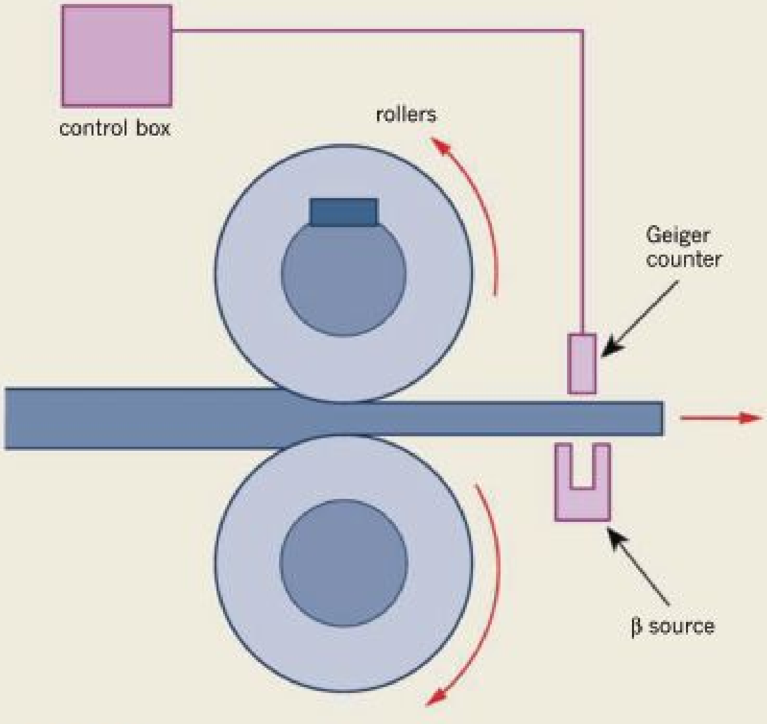


FIGURE 6.3.11 Beta emitters are used to monitor metal-sheet thickness.

PHYSICSFILE

Cobalt bomb

An isotope of cobalt-60 emits γ radiation of very high energy. For this reason it is used to irradiate cancerous cells. The first instruments that used cobalt were placed in a thick lead casing in the shape of a bomb, and so they used to be described as 'cobalt bombs'.

FIGURE 6.3.10 A radiotherapy session with the latest linear accelerator, the Truebeam. A nurse and technician set up a patient to treat head and neck cancer.

PROPERTIES OF α, β AND γ RADIATION

The properties of alpha (α), beta (β) and gamma (γ) radiation are summarised in Table 6.3.1.

Property	α particle	β particle	γ-ray
mass	heavy	light	none
speed	up to 20000 km s^{-1} or about 10% of the speed of light	about 90% of the speed of light	the speed of light
charge	+2	−1 or +1	0
range in air	a few centimetres	1 or 2 m	many metres
penetration in matter	~10^{-2} mm	a few mm	high

TABLE 6.3.1 A comparison of the properties of alpha, beta and gamma radiation.

EXTENSION

Ionising radiation

Some types of radiation, such as radio waves, are harmless. Other types, however, are dangerous to humans. Known as ionising radiation, these harmful types of radiation interact with atoms. They have enough energy to remove outer-shell electrons and create ions. Alpha particles, beta particles and gamma rays are all ionising. So too is electromagnetic radiation with a frequency above 2×10^{16} Hz. So, X-rays and ultraviolet-B radiation are also ionising. When ionising radiation interacts with human tissue, the ions it produces are harmful and can lead to the development of cancers and tumours. Of the three types of radiation, alpha has the greatest ionising power, followed by beta and then gamma rays.

Lower energy electromagnetic radiation, such as radio waves, microwaves, infrared, visible light and ultraviolet-A radiation are non-ionising. Exposure to this radiation, even in significant amounts, has no serious consequences. Non-ionising radiation does not have enough energy to change the chemistry of the atoms and molecules that make up body cells.

6.3 Review

SUMMARY

- Alpha (α) particles are ejected from the nucleus at around 10% of the speed of light. They have a double positive charge and are relatively heavy. Alpha particles have poor penetrating power.
- Beta (β) particles emanate from the nucleus at up to 90% of the speed of light. They are much lighter than α particles. β^- particles have a single negative charge. β^+ particles have a single positive charge. Beta radiation has a moderate penetrating ability.
- Gamma (γ) rays are high-energy electromagnetic radiation and so travel at the speed of light and have no charge. They have high penetrating power.

KEY QUESTIONS

1 Which type of radiation (alpha, beta or gamma):
  a can easily penetrate aluminium foil?
  b is ejected when a neutron decays into a proton?
  c travels relatively slowly at typically around 10% of the speed of light?
  d travels at speeds of up to 90% of the speed of light?
  e has no charge?

2 Which type of radiation (alpha, beta or gamma) is unaffected by a magnetic field?

3 Which type of radiation (alpha, beta or gamma) could penetrate human skin but not 1 mm of aluminium?

4 Which type of radiation (alpha, beta or gamma) would be used to treat brain cancer, where the radiation needs to penetrate the skull and reach the site of the tumour?

5 A radioactive source is emitting alpha, beta and gamma radiation into the air. Which type/s of radiation would a Geiger counter held about 20 cm from the source most likely detect?

6 Where in the atom does each type of radiation originate from?
  a alpha
  b beta
  c gamma

7 List the radiations alpha, beta and gamma in order of decreasing penetrating power.

8 Briefly explain why alpha particles have a very poor penetrating ability.

9 A radiographer inserts a radioactive wire into a breast cancer to destroy the cancerous cells close to the wire. Should this wire be an alpha, beta or gamma emitter? Explain your reasoning.

10 Explain why beta particles, not alpha or gamma, are the best to use for monitoring the thickness of thin metal sheets.

6.4 Half-life and decay series

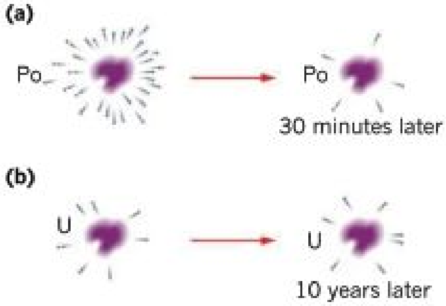


FIGURE 6.4.1 The activity of (a) polonium-218 and (b) uranium-235.

Different radioisotopes will emit radiation and will decay at very different rates. A Geiger counter held close to a small sample of polonium-218 will initially detect a very high level of radiation. Half an hour later, this count rate will have dropped to almost zero.

Compare this with a similar sample of uranium-235. A Geiger counter held close to the uranium will show a low count rate. However, as time passes, the count rate does not seem to change. If you came back decades later, the count level would still be the same. Figure 6.4.1 compares the activity of both of these radioisotopes.

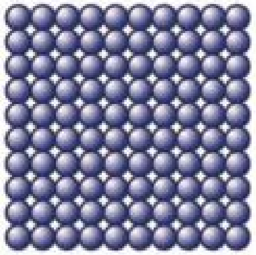
The **half-life** of a radioisotope describes how long it takes for half of the atoms in a given mass to decay. The count rate is the **activity** of the sample. These ideas will be studied in this section.

HALF-LIFE

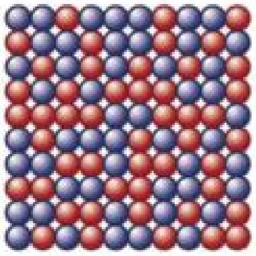
All radioisotopes are unstable but some are more unstable than others. In the previous example, polonium-218 is more unstable than uranium-235. One way of determining this instability is by measuring the half-life ($t_{1/2}$) of the radioisotope.

The half-life ($t_{1/2}$) of a radioisotope is the time that it takes for half of the nuclei of the sample radioisotope to decay.

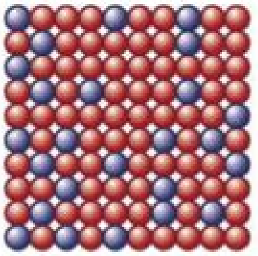
Key: ● = 1 million ^{218}Po nuclei



Initially:
100 million ^{218}Po nuclei



After 3 minutes:
~ 50 million ^{218}Po nuclei



After 6 minutes:
~ 25 million ^{218}Po nuclei

FIGURE 6.4.2 The decay of polonium-218 over two half-lives, showing how only one quarter (25%) of the original radioisotope remains after two half-lives.

The half-life of polonium-218 is 3 minutes. Consider a sample of polonium that initially contains 100 million undecayed polonium-218 nuclei, as shown in Figure 6.4.2. Over the first 3 minutes about half of these will have decayed, leaving around 50 million polonium-218 nuclei. Over the next 3 minutes half of these 50 million polonium-218 nuclei will decay, leaving approximately 25 million of the original radioactive nuclei. The process continues as time passes.

The number of nuclei remaining after a particular number of half-lives can be found mathematically using:

$$N = N_0 \left(\frac{1}{2}\right)^n$$

where N is the number of radioactive nuclei remaining

N_0 is the initial number of radioactive nuclei

n is the number of half-lives elapsed.

The number of half-lives in a period of time can be found using:

$$n = \frac{T}{t_{1/2}}$$

where n is the number of half-lives elapsed

T is the period of time that the radioactive nuclei has decayed

$t_{1/2}$ is the half-life of the radioactive nuclei.

As time passes, a smaller and smaller proportion of the original radioisotope remains in the sample. The graph in Figure 6.4.3, known as a decay curve, shows this.

Even a very small radioactive sample will contain billions of atoms. It is important to know that although the behaviour of such a large sample of nuclei can be mathematically predicted, the behaviour of one particular nucleus cannot. It has a 50% chance of decaying in each half-life. Also, the half-life of a radioisotope is constant and cannot be changed by chemical reactions, heat and so on. Half-life is solely determined by the instability of the nuclei of the radioisotope.

Worked example 6.4.1

HALF-LIFE

A sample of the radioisotope thorium-234 contains 8.0×10^{12} nuclei. The half-life of thorium-234 is 24 days. How many thorium-234 nuclei will remain in the sample after 120 days?

Thinking	Working
Calculate how many half-lives 120 days corresponds to.	$n = \frac{120}{24}$ $= 5$ half-lives
Substitute $N_0 = 8.0 \times 10^{12}$ and $n = 5$ into the equation. Calculate the number of nuclei remaining.	$N = N_0\left(\frac{1}{2}\right)^n$ $= 8.0 \times 10^{12} \times \left(\frac{1}{2}\right)^5$ $= 2.5 \times 10^{11}$ nuclei

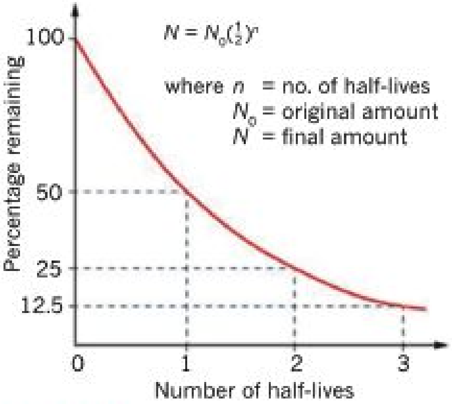


FIGURE 6.4.3 A decay curve for a radioisotope.

Worked example: Try yourself 6.4.1

HALF-LIFE

A sample of the radioisotope sodium-24 contains 4.0×10^{10} nuclei. The half-life of sodium-24 is 15 hours. How many sodium-24 atoms will remain in the sample after 150 hours?

ACTIVITY

A Geiger counter can be used to record the number of radioactive decays occurring in a sample each second. It measures the activity of the sample.

Activity is measured in becquerels, Bq.

1 Bq = 1 disintegration per second.

Over time, the activity of a sample of a radioisotope will decrease. More and more of the radioactive nuclei will have decayed and at some point it will no longer emit radiation.

If a sample of polonium-218 ($t_{1/2} = 3$ minutes) has an initial activity of 2000 Bq, then after one half-life its activity will be 1000 Bq. After a further 3 minutes, the activity of the sample will have reduced to 500 Bq, and after a further 3 minutes it will be 250 Bq and so on.

Uranium-235 has a half-life of 700 000 years. Its activity will remain virtually constant for decades and will certainly not change over 3 minutes.

The same half-life equation can be used to calculate the final activity of a radioactive sample after a number of half-lives.

The activity of the nuclei remaining after a number of half-lives can be found mathematically using:

$$A = A_0\left(\frac{1}{2}\right)^n$$

where A is the activity of radioactive nuclei remaining

A_0 is the initial activity of radioactive nuclei

n is the number of half-lives elapsed.

COMMON RADIOISOTOPES AND THEIR APPLICATIONS

The half-lives of some common radioisotopes are shown in Table 6.4.1. The half-life of a radioisotope will determine what it is used for. For example, the most commonly used medical tracer, technecium-99, has a short half-life of just 6 hours. The short half-life means that radioactivity does not remain in the body any longer than necessary. On the other hand, the radioisotope used in a smoke detector, americium-241, is chosen because of its long half-life, 461 years. The smoke detector can continue to function for a very long time, as long as the battery is replaced each year.

Isotope	Emission	Half-life	Application
Natural			
polonium-214	α	0.000 16 s	nothing at this time
strontium-90	β	28.8 years	cancer therapy
radium-226	α	1630 years	once used in luminous paints
carbon-14	β	5730 years	carbon dating of fossils
uranium-235	α	700 000 years	nuclear fuel, rock dating
uranium-238	α	4.5 billion years	nuclear fuel, rock dating
thorium-232	α	14 billion years	fossil dating, nuclear fuel
Artificial			
technetium-99m	β	6 h	medical tracer
sodium-24	β	15 h	medical tracer
iodine-131	β	8 days	medical tracer
phosphorus-32	β	14.3 days	medical tracer
cobalt-60	β	5.3 years	radiation therapy
americium-241	α	460 years	smoke detectors
plutonium-239	α	24 000 years	nuclear fuel, rock dating

TABLE 6.4.1 Some common radioisotopes and their half-lives.

DECAY SERIES

Generally, when a radionuclide decays, its daughter nucleus is not completely stable and is itself radioactive. This daughter nucleus will then undergo further decay. Eventually a stable isotope is reached and the sequence ends. This is known as a **decay series**. An example of a decay series is shown in Figure 6.4.4. This particular series shows the decay of uranium-238 (shown at the top of the chart) into lead-206 (shown at the bottom of the chart).

The Earth is 4.5 billion years old—old enough to have only four naturally occurring decay series that remain active. These are:

- the uranium series in which uranium-238 eventually becomes lead-206
- the actinium series in which actinum-235 eventually becomes lead-207
- the thorium series in which thorium-232 eventually becomes lead-208
- the neptunium series in which neptunium-237 eventually becomes bismuth-209. (Since neptunium-237 has a relatively short half-life, it is no longer present in the crust of the Earth, but the rest of its decay series is still continuing.)

Geologists analyse the proportions of the radioactive elements in a sample of rock to gain a reasonable estimate of the rock's age. This technique is known as rock dating.

FIGURE 6.4.4 The uranium decay series. The half-life and emissions are indicated on each of the decays as radioactive uranium-238 is gradually transformed into stable lead-206. Mining companies find significant quantities of lead at uranium mines.

EXTENSION

Radiocarbon dating

Carbon dating is a technique used by archaeologists to determine the ages of fossils and ancient objects that were made from plant matter. This method involves measuring and comparing the proportion of two isotopes of carbon, ^{12}C and ^{14}C, in the specimen.

Carbon-12 is a stable isotope but carbon-14 is radioactive. Carbon-14 only exists in trace amounts in nature. Carbon-12 atoms are about 1 000 000 000 000 (10^{12}) times more common than carbon-14 atoms.

Carbon-14 has a half-life of 5730 years and decays by β^- emission to nitrogen-14. Its decay equation is:

$$^{14}_{6}C \rightarrow {}^{14}_{7}N + \beta^- + \bar{\nu}$$

Both carbon-12 and carbon-14 can combine with other atoms in the environment. For example, they both combine with oxygen to form carbon dioxide. Plants and animals take in carbon-based molecules from the air and food. This means that all living organisms contain the same percentage of carbon-14. In the environment, the production of carbon-14 is matched by its decay and so the proportion of carbon-14 atoms to carbon-12 remains constant.

After an organism dies, the amount of carbon-14 it contains will decrease as these atoms decay to form nitrogen-14 and are not replaced from the environment. The number of atoms of carbon-12 does not change as carbon-12 is a stable atom. So, over time, the proportion of carbon-14 to carbon-12 atoms decreases.

The proportion of carbon-14 to carbon-12 in a dead organism can be compared with that found in living organisms and the approximate age of the specimen can be determined from the half-life of carbon-14.

Consider this example. The count rate from a 1 gram sample of carbon that has been extracted from an ancient wooden spear is 10 Bq. A 1 gram sample of carbon from a living piece of wood has a count rate of 40 Bq. We can assume that this was also the initial count rate of the spear. For its count rate to have reduced from 40 to 10 Bq, the spear must be (40 → 20 → 10) two half-lives of carbon-14 old, i.e. about 11 500 years old.

In 1988, scientists used carbon-dating techniques to show that the Shroud of Turin was probably a medieval forgery. It had been claimed that the Shroud of Turin was the piece of cloth that was the burial shroud of Jesus Christ (see Figure 6.4.5). Carbon-dating tests on small samples of the cloth established that there was a high probability that it was made between 1260 and 1390, not around the time of Christ's death.

FIGURE 6.4.5 The Shroud of Turin.

Radiocarbon dating is an important aid to anthropologists who are interested in finding out about the migration patterns of early people—including the Australian Aborigines. This technique is very powerful since it can be applied to the remains of ancient campfires. It is accurate and reliable for samples up to about 60 000 years old. Carbon dating cannot be used to date dinosaur bones as they are more than 60 million years old, but it can be used to determine the age of more recently extinct mammoth fossils, like that shown in Figure 6.4.6.

FIGURE 6.4.6 A fossilised mammoth analysed by carbon dating.

6.4 Review

SUMMARY

- The rate of decay of a radioisotope is measured by its half-life $t_{1/2}$. This is the time that it takes for half of the radioisotope to decay.
- The activity of a sample indicates the number of emissions per second. Activity is measured in becquerels (Bq), where 1 Bq = 1 emission per second.
- The number of atoms of a radioisotope will decrease over time. Over one half-life, the number of atoms of a radioisotope will halve.
- The half-life equation can be used to calculate the number (N) or activity (A) of a radioisotope remaining after a number of half-lives (n) has passed:
- $N = N_0\left(\frac{1}{2}\right)^n$, $A = A_0\left(\frac{1}{2}\right)^n$
- When a radionuclide decays, its daughter nucleus is usually itself radioactive. This daughter will then decay to a grand-daughter nucleus, which may also be radioactive, and so on. This is called a decay series.

KEY QUESTIONS

1 What is meant by the 'activity' of a radioisotope?

2 Technetium-99m has a half-life of 6.0 hours. A sample of the radioisotope originally contains 8.0×10^{10} atoms. How many technetium-99m nuclei remain after 6.0 hours?

3 Iodine-131 has a half-life of 8 days. A sample of the radioisotope initially contains 2.4×10^{12} iodine-131 nuclei. How many iodine-131 nuclei remain after 24 days?

4 Radioactive materials are considered to be relatively safe when their activity has fallen below 0.1% of the initial value.

a How many half-lives does this take?

b Plutonium-239 is a by-product of nuclear reactors. Its half-life is 24 000 years. How long does the plutonium-239 have to be stored as nuclear waste before it is considered safe to handle?

5 If a particular atom in a radioactive sample has not decayed during the previous half-life, what is the percentage chance that it will decay in the next half-life?

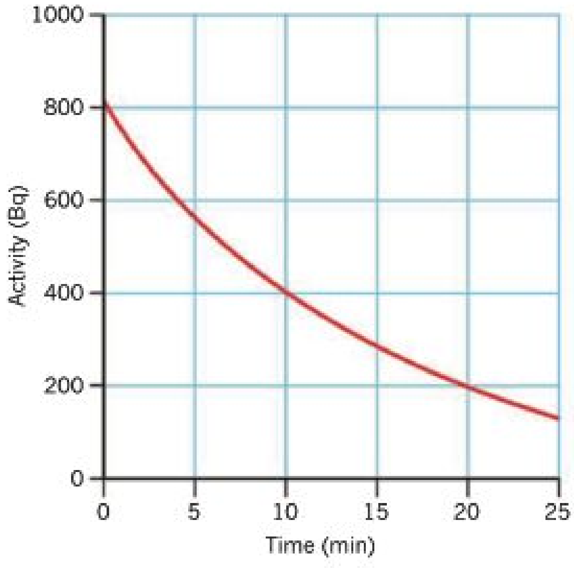
6 A hospital in Alice Springs needs 12 μg of the radioisotope technetium-99m. The specimen has to be ordered from Sydney. The half-life of technetium-99m is 6 hours and the delivery takes 24 hours. How much must be produced in Sydney to satisfy the Alice Springs order?

7 The activity of a radioisotope changes from 6000 Bq to 375 Bq over a period of 60 minutes. Calculate the half-life of this radioisotope.

8 A Geiger counter is used to measure the radioactive emissions from a certain radioisotope. The activity of the sample is shown in the graph.

a What is the half-life of the radioisotope according to the graph?

b What would the activity of the sample be after 40 minutes have elapsed?



9 According to Figure 6.4.4 on page 220, what type of decay does lead-210 undergo and what is its half-life?

10 In the uranium decay series shown in Figure 6.4.4, $^{238}_{92}$U decays to eventually produce stable $^{206}_{82}$Pb. How many alpha and beta-minus decays have occurred?

Chapter review

KEY TERMS

activity
alpha particle
antineutrino
atomic number
beta particle
daughter nucleus
decay series
electron
electrostatic force
gamma ray
Geiger counter
half-life
isotope
mass number
neutral
neutron
nuclear transmutation
nucleon
nucleus
nuclide
parent nucleus
penetrating ability
positron
proton
radiation
radioactive
radioisotope

1 How many protons and neutrons are in the $^{41}_{20}\text{Ca}$ nuclide?

2 Use the periodic table in Figure 6.1.7 on page 203 to determine the number of protons, neutrons and nucleons in cobalt-60.

3 Determine the nature of the unknown, X, for the following transmutation:

$^{60m}_{27}\text{Co} \rightarrow ^{60}_{27}\text{Co} + \text{X}$

(60m means the nuclide is metastable and has a higher level of stability than very short-lived isotopes. The mass number is still 60.)

4 What type of radiation does potassium-48 (atomic number 19) emit? Use Figure 6.2.6 on page 208 to answer this question.

5 Identify each of these radiation types:

- **a** $^{0}_{-1}\text{A}$
- **b** $^{1}_{1}\text{B}$
- **c** $^{4}_{2}\text{C}$
- **d** $^{1}_{0}\text{D}$
- **e** $^{0}_{0}\text{E}$
- **f** $^{0}_{1}\text{F}$

6 Some nuclei can be made unstable by firing neutrons into them. The neutron is captured and the nucleus becomes unstable. The nuclear equation when the stable isotope boron-10 transmutates by neutron capture into a different element, X, by emitting alpha particles is:

$^{10}_{5}\text{B} + ^{1}_{0}\text{n} \rightarrow \text{X} + ^{4}_{2}\text{He}$

Identify the unknown element, X, and its mass and atomic numbers.

7 Identify each of the unknown particles X and Y in the following nuclear transmutations.

- **a** $^{14}_{7}\text{N} + \alpha \rightarrow ^{17}_{8}\text{O} + \text{X}$
- **b** $^{27}_{13}\text{Al} + \text{Y} \rightarrow ^{27}_{12}\text{Mg} + ^{1}_{1}\text{H}$

8 Find the values of *x* and *y* in each of these radioactive decay equations.

- **a** $^{208}_{81}\text{Ti} \rightarrow ^{x}_{y}\text{Pb} + \beta^-$
- **b** $^{180}_{80}\text{Hg} \rightarrow ^{x}_{y}\text{Pt} + \alpha$

9 Fluorine-18 is a radioisotope that is used for detecting tumours. It is formed when radioactive neon-18 decays by positron emission. Fluorine-18 in turn also decays by positron emission. The equations are as follows:

$^{18}_{10}\text{Ne} \rightarrow ^{a}_{b}\text{X} + ^{0}_{+1}\beta$

$^{a}_{b}\text{X} \rightarrow ^{c}_{d}\text{Y} + ^{0}_{+1}\beta$

Determine the values of *a*, *b*, *c*, *d* and identify X and Y, which are the daughter nuclei that result from this process.

10 The radioisotope nitrogen-12 decays by emitting a positron and a neutrino. The decay equation for nitrogen-12 is:

$^{12}_{7}\text{N} \rightarrow \text{X} + ^{0}_{+1}\beta + ^{0}_{0}\nu$

Identify particle X.

11 A stable isotope of neon has 10 protons and 10 neutrons in each nucleus. Every proton is repelling all the other protons. Why is the nucleus stable?

12 Which type of radiation out of alpha, beta and gamma:

- **a** is the fastest?
- **b** has the greatest penetrating power?

13 Health workers who deal with radiation to treat cancer often have to wear a lead vest to protect their vital organs from exposure. Which type(s) of radiation is the lead apron shielding them from?

14 A nuclear physicist was bombarding a sample of beryllium-7 with a beam of electrons in an effort to smash the electrons into the beryllium nuclei. Why would it be quite difficult for a collision between the electrons and the nuclei to occur?

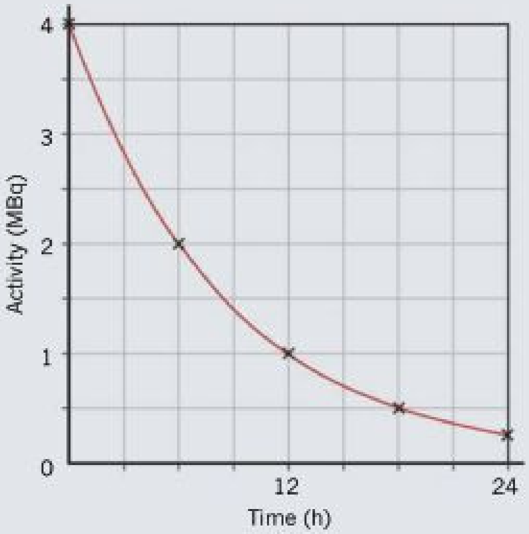
15 A radioactive isotope X has a half-life of 20 minutes. A sample starts with 6.0×10^{14} atoms of the isotope. What amount of the original isotope will remain after 20 minutes?

16 Radioisotope Y has a half-life of 3.0 hours. A sample starts with 5.6×10^{15} atoms of the radioisotope. How many atoms of Y remain after 9.0 hours?

Chapter review *continued*

17 Uranium-235 has a half-life of 700 000 000 years, while uranium-238 has a half-life many times longer of 4.5×10^9 years. In samples of 1 kg of each of these pure radioisotopes, which one would have the greater activity?

18 The decay curve for a sample of the radioisotope technetium-99m is shown. It has an initial activity of 4.0×10^6 Bq.



a What is the activity of this sample after one half-life?

b From the graph, what is the half-life of technetium-99m?

c If the sample is produced in a hospital at 4 pm, what is its activity when it is used at 10 am the next day?

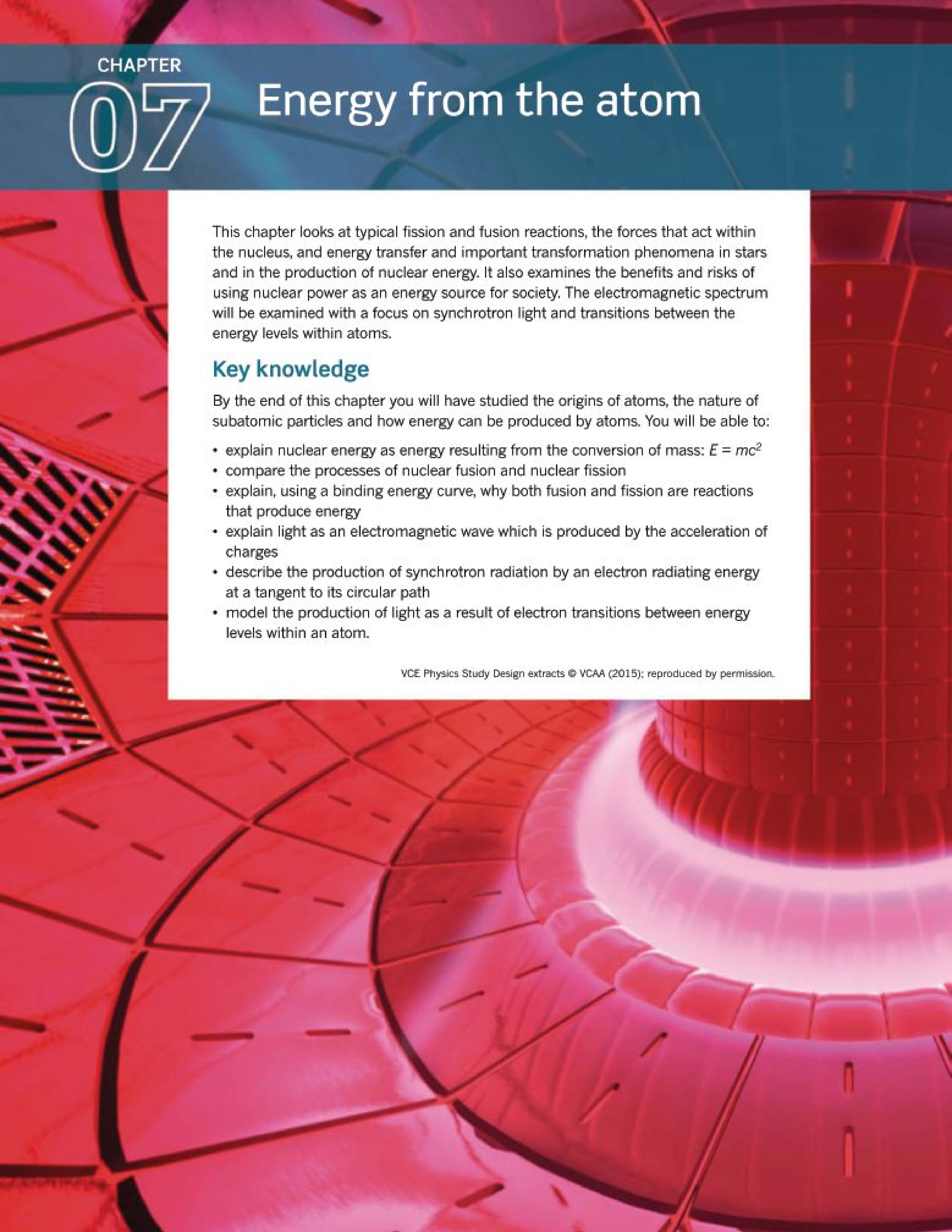
19 Protactinium-234 is a radioactive element with a half-life of 70 s. If a sample of this radioisotope contains 6.0×10^{10} nuclei, how many nuclei of this element will remain after 140 s?

20 Radiotherapy treatment of brain tumours involves irradiating the target area with radiation from an external source. Why is cobalt-60 (gamma emitter with a half-life of 5.3 years) generally used as the radiation source for this treatment?

CHAPTER

07

Energy from the atom



This chapter looks at typical fission and fusion reactions, the forces that act within the nucleus, and energy transfer and important transformation phenomena in stars and in the production of nuclear energy. It also examines the benefits and risks of using nuclear power as an energy source for society. The electromagnetic spectrum will be examined with a focus on synchrotron light and transitions between the energy levels within atoms.

Key knowledge

By the end of this chapter you will have studied the origins of atoms, the nature of subatomic particles and how energy can be produced by atoms. You will be able to:

- explain nuclear energy as energy resulting from the conversion of mass: $E = mc^2$
- compare the processes of nuclear fusion and nuclear fission
- explain, using a binding energy curve, why both fusion and fission are reactions that produce energy
- explain light as an electromagnetic wave which is produced by the acceleration of charges
- describe the production of synchrotron radiation by an electron radiating energy at a tangent to its circular path
- model the production of light as a result of electron transitions between energy levels within an atom.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

7.1 Nuclear fission and energy

In 1905, Albert Einstein theorised that mass, m, and energy, E, are equivalent through the equation $E = mc^2$. This led to the realisation that vast amounts of energy lie unharnessed within the nuclei of atoms. The ramifications of Einstein's work and the discovery of nuclear fission were realised in 1945 with the explosion of the first atomic bomb in the desert near Alamogordo in New Mexico, USA (see Figure 7.1.1). In this section, nuclear fission and the energy that it can unleash will be explored.

FIGURE 7.1.1 An atomic bomb explosion and its associated mushroom cloud.

INSIDE THE NUCLEUS

The current understanding of the basic properties and structure of the nucleus is the result of intense scientific investigation in the early part of the twentieth century. Physicists such as Becquerel, Rutherford, Chadwick, Geiger, Marsden and Harkins were instrumental in the development of the model of the nucleus that exists today. These renowned scientists are shown in Figure 7.1.2.

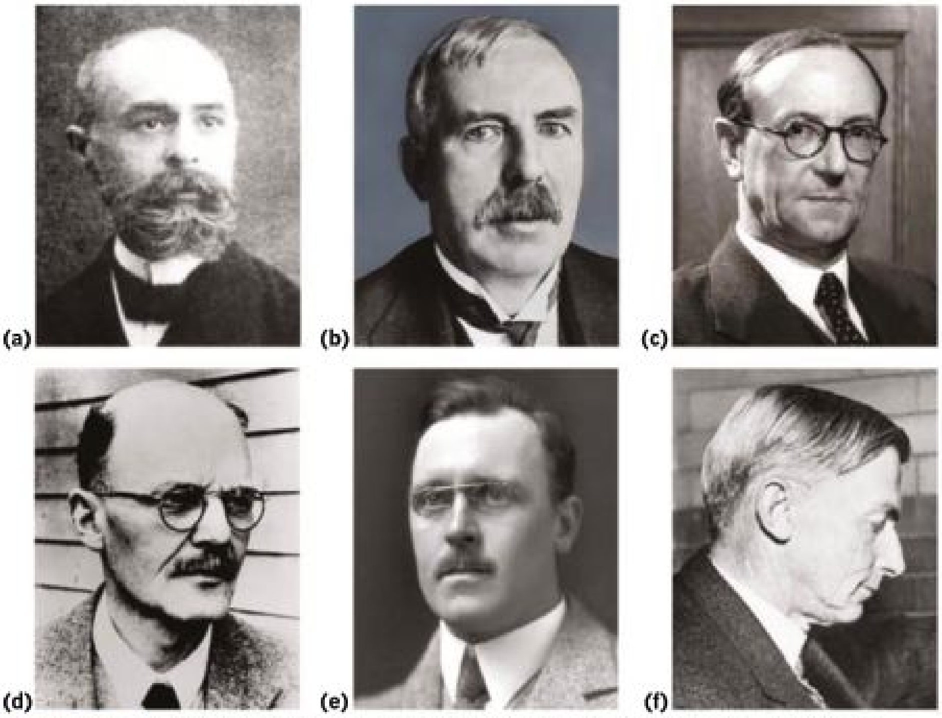


FIGURE 7.1.2 (a) Henri Becquerel, (b) Ernest Rutherford, (c) James Chadwick, (d) Hans Geiger, (e) Ernest Marsden and (f) William Harkins.

Recall that within the nucleus, there are protons in close proximity to other protons. This should seem odd, since protons exert strong electrostatic forces of repulsion on each other. In other words, protons are repelled by other protons in the nucleus. The nucleus should simply blow apart. It doesn't, so there must be something else going on. Another force, known as the strong nuclear force, is also acting, and this is a force of attraction that acts to hold the nucleus together.

Electrostatic forces act between charged particles and can act over relatively large distances. Within a nucleus, this means that each proton is strongly repelling every other proton, so this force is trying to make the nucleus disintegrate. Neutrons are unaffected by electrostatic forces.

The strong nuclear force is a force of attraction that acts between every nucleon. This force acts like a nuclear cement. Neutrons are attracted to nearby neutrons and protons. Protons are similarly attracted to nearby neutrons and protons. However, this force only acts over relatively short distances. For nucleons on opposite sides of a large nucleus, this force is not significant.

An example of how the electrostatic and strong nuclear forces act in a nucleus is shown in Figure 7.1.3.

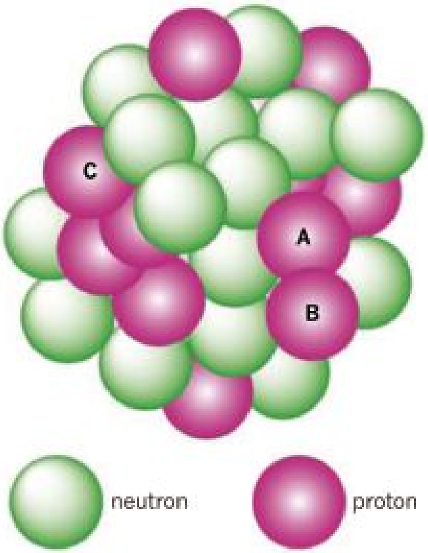


FIGURE 7.1.3 Proton A both attracts and repels proton B, but, at short distances, the attraction due to the strong nuclear force is much greater than the repulsion due to the electrostatic force. Proton A also both attracts and repels proton C, but because of the greater distance between them, the force of repulsion is larger. However, proton A and proton C do not fly apart due to the strong attractive forces exerted on them by adjacent neutrons.

PHYSICSFILE

Strong nuclear force

The existence of the strong nuclear force was first proposed by Japanese theoretical physicist Hideki Yukawa in 1935. However, the properties of this force are so complex that it took until 1975 for physicists to develop a mathematical model that could successfully describe it.

NUCLEAR FISSION

The discovery of the neutron by James Chadwick in 1932 enabled scientists to explore the behaviour of larger atomic nuclei. Up until then, physicists such as Enrico Fermi had been firing α-particles at target nuclei and analysing the results. Chadwick found that with larger target nuclei, the positive α-particles were too strongly repelled from the positively charged nuclei and collisions did not occur.

The advantage of a neutron is that it is neutral and so is not repelled by any target nucleus. The bombarding neutrons can be absorbed into the nucleus of the target atom, as shown in Figure 7.1.4. This makes neutrons very useful as a form of radiation. They are used in many experiments to artificially transmutate (change the form of) different isotopes.

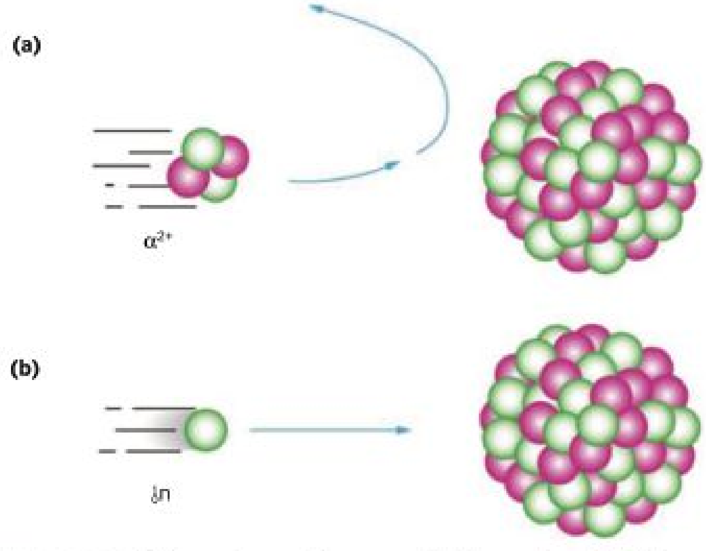


FIGURE 7.1.4 (a) Charged α-particles are repelled by a nucleus. (b) Uncharged neutrons are able to smash into a nucleus.

Nuclear **fission** occurs when an atomic nucleus splits into two or more pieces. This is usually triggered or induced by the absorption of a neutron, as shown in Figure 7.1.5. Nuclides that are capable of undergoing nuclear fission after absorbing a neutron are said to be **fissile**. Only a handful of fissile nuclides exist in nature.

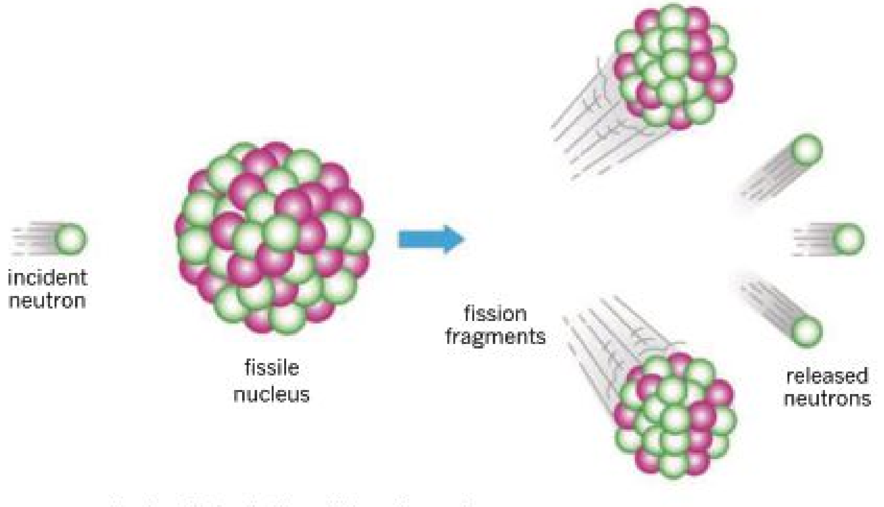


FIGURE 7.1.5 Nuclear fission is the splitting of a nucleus.

Uranium-235 and plutonium-239 are fissile and can be made to split when bombarded by a slow-moving neutron. Uranium-238 and thorium-232 require a very high-energy neutron to induce fission, so they are regarded as fissionable, but non-fissile.

RELEASE OF NEUTRONS DURING FISSION

Uranium-235 and plutonium-239 are the fissile nuclides most commonly used in nuclear reactors and nuclear weapons. They are more fissile than uranium-238 and thorium-232.

When a uranium-235 or plutonium-239 nucleus absorbs either a slow- or fast-moving neutron, it becomes unstable and spontaneously undergoes fission. However, fission is more likely to be induced by a slow-moving neutron because it is more easily captured.

A uranium-235 nucleus may split in many different ways. When a sample of uranium-235 undergoes fission, a wide variety of fission products are produced. Figure 7.1.6 shows one outcome but many others are possible. Usually either two or three neutrons are released. For uranium-235, an average of 2.47 neutrons per fission has been determined.

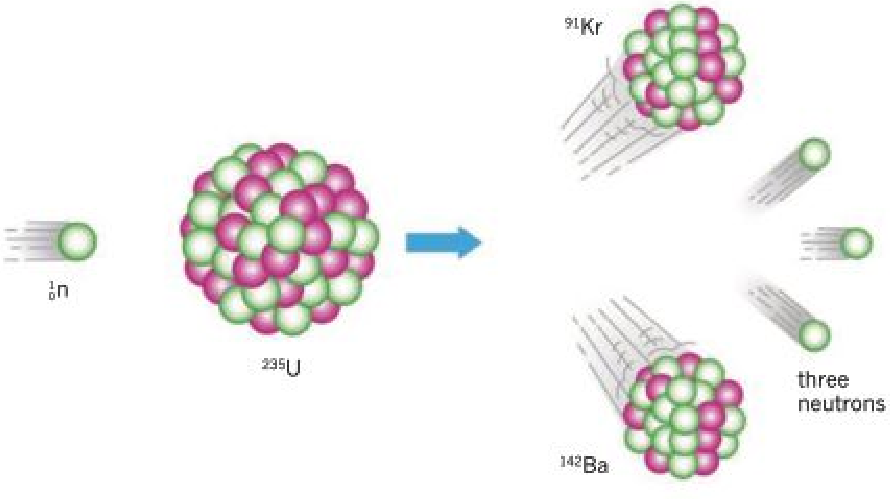


FIGURE 7.1.6 One possible outcome for the neutron-induced fission of uranium-235. This example shows three neutrons released along with krypton-91 and barium-142 daughter nuclides.

A typical fission reaction for uranium-235 is:

$${}^{1}_{0}\mathrm{n} + {}^{235}_{92}\mathrm{U} \rightarrow {}^{236}_{92}\mathrm{U} \rightarrow {}^{91}_{36}\mathrm{Kr} + {}^{142}_{56}\mathrm{Ba} + 3{}^{1}_{0}\mathrm{n} + \text{energy}$$

Krypton-91 and barium-142 are known as the daughter nuclei or **fission fragments**. Three neutrons are freed from this uranium nucleus when it splits. Note that in the same way as for radioactive decay, both the atomic number, Z, and mass number, A, are conserved in these nuclear reactions. For the reaction equation shown, the atomic numbers on either sides of the arrows add up to 92 and the mass numbers add up to 236.

In the end, the decay products of the nuclear fission process form a lethal cocktail of radioactive isotopes. It is these radioactive fission fragments that comprise the bulk of the high-level waste produced by nuclear reactors.

Plutonium-239 will also undergo fission in a variety of ways. It releases an average of 2.89 neutrons per fission, slightly more than uranium-235.

HOW TO MEASURE ENERGY IN ELECTRONVOLTS

The energy of moving objects such as cars and tennis balls is measured in joules. However, nuclei, subatomic particles and radioactive emissions have such small amounts of energy that the joule is inappropriate.

The energy of subatomic particles and radiation is usually given in **electronvolts** (eV). One electronvolt is an extremely small amount of energy.

An electronvolt is the energy that an electron would gain if it were accelerated by a voltage of 1 volt and is equal to 1.6×10^{-19} J.

To convert from eV to joules: multiply by 1.6×10^{-19} J.

To convert from joules to eV: divide by 1.6×10^{-19} J.

ENERGY RELEASED DURING NUCLEAR REACTIONS

It is well established that the mass of any nucleus is always less than the mass of its individual nucleons. Two separate protons and two separate neutrons will have slightly more total mass than a helium nucleus.

Albert Einstein, pictured in Figure 7.1.7, provided the explanation of the origins of this missing mass. He showed that *mass* and *energy* were not completely independent quantities. Indeed, mass can be converted into energy and energy can be converted into mass.

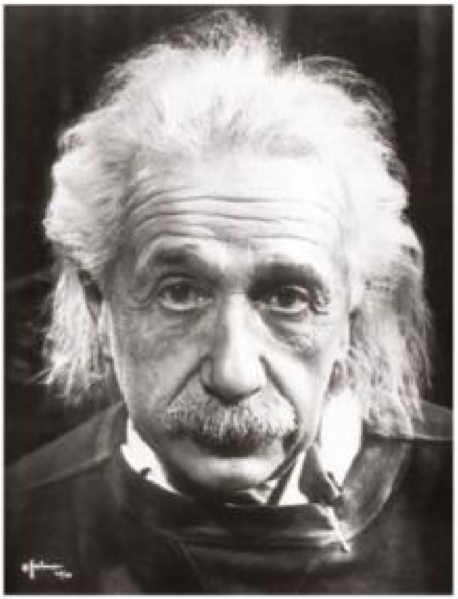


FIGURE 7.1.7 Albert Einstein.

If you wanted to separate a helium nucleus into four free nucleons, you would need to add energy to the nucleus. This energy is the **binding energy** of the nucleus. The free nucleons will have more energy and so, according to Einstein, will have greater mass.

The energy released as a result of a mass defect (mass decrease) is given by Einstein's famous equation:

$E = mc^2$

where E is energy (J)

m is the mass defect (the decrease in mass, in kg)

c is the speed of light = 3.0×10^8 m s^{-1}.

The chemical reactions that you have probably performed at school typically release only a few electronvolts of energy. Compared with this, an enormous amount of energy is released during nuclear reactions. This has made nuclear energy the focus of scientific research over the past century.

During radioactive decay millions of electronvolts of energy can be released. Alpha particle decay usually involves the release of 5–10 MeV (5–10 million electronvolts) of energy. Nuclear fission involves much more energy again, typically around 200 MeV. This energy is mainly in the form of the kinetic energy of the fission fragments and neutrons, as well as the emission of energy as gamma radiation.

During any fission reaction, the combined mass of the incident neutron and the target nucleus is always slightly greater than the combined mass of the fission fragments and the released neutrons. For example, in Figure 7.1.8 (p. 230), the mass of the incident neutron and the uranium-235 nucleus is greater than the combined masses of the fission products—barium-142, krypton-91 and the three neutrons.

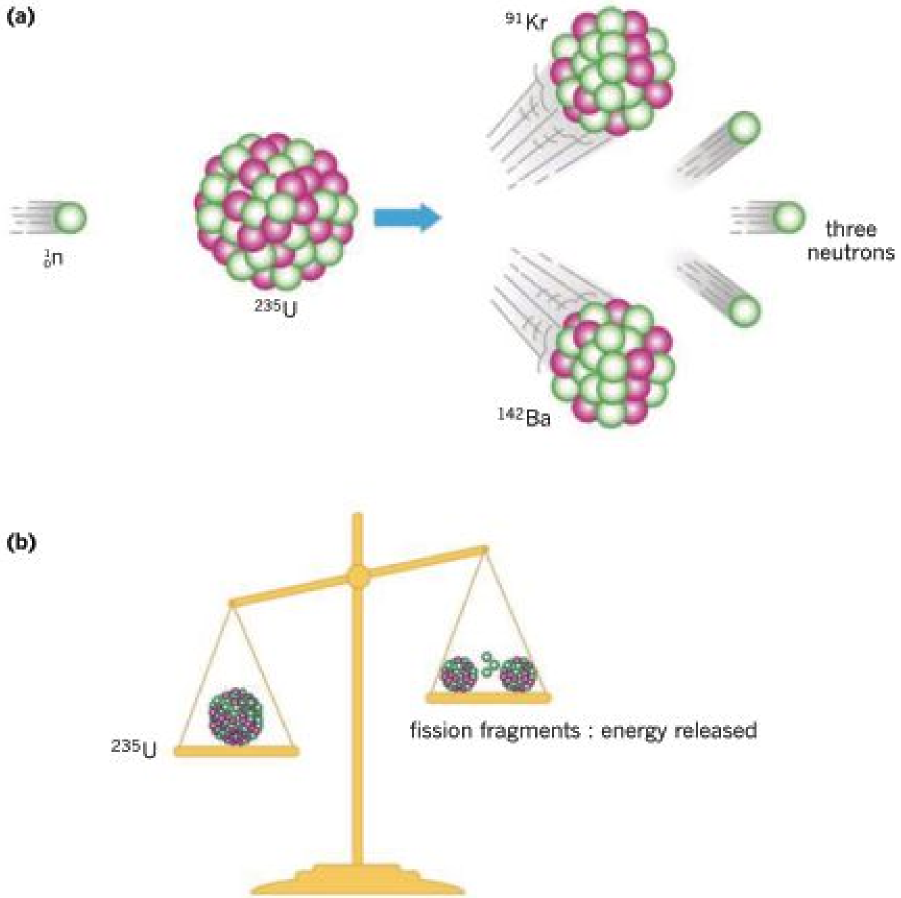


FIGURE 7.1.8 (a) Each fission of a uranium-235 nucleus releases about 200 MeV of energy. (b) The fission fragments have a lower combined mass than the uranium nucleus. The missing mass is converted into the 200 MeV of energy according to the equation $E = mc^2$.

Only a very small proportion of the original mass of the nuclei is available as usable energy—typically around 0.1%. If you had a 1.000 kg block of pure uranium-235 that underwent fission completely, at the end you would have a block of radioactive fission fragments with a mass of around 0.999 kg.

Worked example 7.1.1

FISSION

Plutonium-239 is a fissile material. When a plutonium-239 nucleus is struck by and absorbs a neutron, it can split in many different ways. Consider the example of a nucleus that splits into barium-145 and strontium-93 and releases some neutrons.

The nuclear equation for this is:

$^{1}_{0}n + ^{239}_{94}Pu \rightarrow ^{145}_{56}Ba + ^{93}_{38}Sr + a\,^{1}_{0}n + \text{energy}$

a How many neutrons are released during this fission process, i.e. what is the value of *a*?

Thinking	Working
Analyse the mass numbers (*Z*).	$1 + 239 = 145 + 93 + (a \times 1)$ $a = (1 + 239) - (145 + 93)$ $= 2$ 2 neutrons are released during fission.

b During this single fission reaction, there was a loss of mass (a mass defect) of 3.07×10^{-28} kg. Calculate the amount of energy that was released during the fission of a single plutonium-239 nucleus. Answer in both MeV and joules.

Thinking	Working
The energy released during the fission of this plutonium nucleus can be found by using $E = mc^2$.	$E = mc^2$ $= (3.07 \times 10^{-28})(3.00 \times 10^8)^2$ $= 2.76 \times 10^{-11}$ J
To convert J into eV, divide by 1.6×10^{-19}. Remember that 1 MeV = 10^6 eV.	$E = \frac{2.76 \times 10^{-11}}{1.6 \times 10^{-19}}$ $= 1.73 \times 10^8$ eV $= 173$ MeV

c The combined mass of the plutonium nucleus and bombarding neutron was 3.99×10^{-25} kg. What percentage of this initial mass was converted into the energy produced during the fission process?

Thinking	Working
Use the relationship mass decrease = $\frac{\text{mass defect}}{\text{initial mass}} \times \frac{100}{1}$.	mass decrease = $\frac{\text{mass defect}}{\text{initial mass}} \times \frac{100}{1}$ $= \frac{3.07 \times 10^{-28}}{3.99 \times 10^{-25}} \times \frac{100}{1}$ $= 0.0769\%$ This is a very small percentage loss in mass.

Worked example: Try yourself 7.7.1

FISSION

Plutonium-239 is a fissile material. When a plutonium-239 nucleus is struck by and absorbs a neutron, it can split in many different ways. Consider the example of a nucleus that splits into lanthanum-143 and rubidium-94 and releases some neutrons.

The nuclear equation for this is:

$${}^{1}_{0}n + {}^{239}_{94}Pu \rightarrow {}^{143}_{57}La + {}^{94}_{37}Rb + a\,{}^{1}_{0}n + \text{energy}$$

a How many neutrons are released during this fission process, i.e. what is the value of *a*?

b During this single fission reaction, there was a loss of mass (a mass defect) of 4.58×10^{-28} kg. Calculate the amount of energy that was released during fission of a single plutonium-239 nucleus. Answer in both MeV and joules. (Give your answers to two significant figures.)

c The combined mass of the plutonium nucleus and bombarding neutron was 2.86×10^{-25} kg. What percentage of this initial mass was converted into the energy produced during the fission process?

PHYSICS IN ACTION

Enrico Fermi, Lise Meitner and nuclear fission

Enrico Fermi, pictured in Figure 7.1.9, was born in Italy in 1901. He completed his doctorate and post-doctorate work in physics at the University of Pisa and in Germany. Fermi had emigrated to the USA by the time the nuclear age dawned in the 1930s. The neutron had just been discovered in 1932, which enabled scientists to fire neutral particles at atomic nuclei for the first time. Fermi was at the forefront of this research.

Fermi bombarded uranium-238 atoms with neutrons and found that uranium-238 nuclei absorbed the neutrons and formed a radioactive isotope of uranium. This isotope then decayed by emitting a beta-minus particle to become neptunium, which then emitted another beta-minus particle to become plutonium, two completely undiscovered elements. Fermi had successfully produced the world's first artificial and **transuranic** (i.e. after uranium) elements. The nuclear reactions for this process are:

$$
\begin{gathered}
{}^{1}_{0}\mathrm{n} + {}^{238}_{92}\mathrm{U} \rightarrow {}^{239}_{92}\mathrm{U} \\
{}^{239}_{92}\mathrm{U} \rightarrow {}^{239}_{93}\mathrm{Np} + {}^{0}_{-1}\beta \\
{}^{239}_{93}\mathrm{Np} \rightarrow {}^{239}_{94}\mathrm{Pu} + {}^{0}_{-1}\beta
\end{gathered}
$$

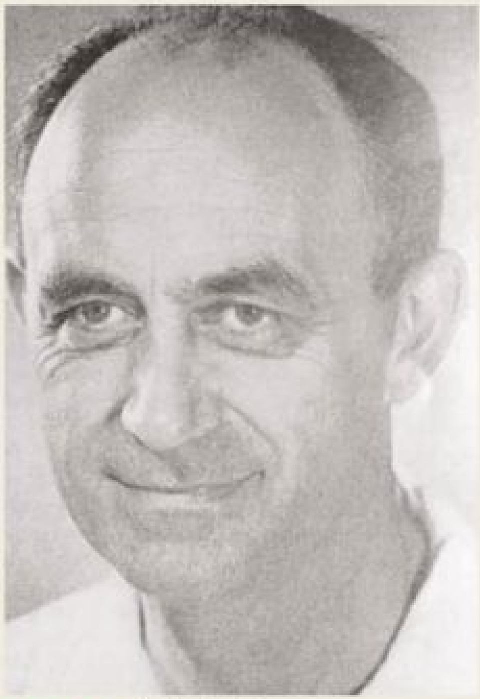


FIGURE 7.1.9 Enrico Fermi.

In 1938, following on from Fermi's work, two German scientists, Otto Hahn and Fritz Strassmann, were also bombarding uranium ($Z = 92$) in an attempt to produce some transuranic elements ($Z > 92$). They found that, rather than producing larger elements, they were getting isotopes of barium ($Z = 56$). Hahn wrote to his colleague Lise Meitner, pictured in Figure 7.1.10, about this unexpected result. She then discussed this with her nephew Otto Frisch, a nuclear physicist, and realised that the bombarding neutrons were causing the uranium nuclei to split. If barium ($Z = 56$) was one of the products, then krypton ($Z = 36$) must be another. This was found to be the case. It was Frisch who coined the term 'fission' and Meitner who proposed that energy would be released during this process.

After the start of World War II, Enrico Fermi was commissioned by President Roosevelt to design and build a device that would sustain the fission process in the form of a chain reaction. In 1942, Fermi succeeded in this task. A squash court at the University of Chicago was used as the site for the world's first nuclear reactor. It produced less than 1 W of power—not even enough to power a small light globe! This sounds like a bit of a failure, but in fact, achieving fission for the first time was a very important breakthrough. The reactor was later modified to produce about 200 W. Fermi died of cancer in 1954. One year after his death, the element with atomic number 100 was artificially produced and named fermium, Fm, in his honour.

FIGURE 7.1.10 Lise Meitner.

7.1 Review

SUMMARY

- Within a nucleus, forces of attraction and repulsion are acting. The long-distance electrostatic force of repulsion acts between the protons. A short-distance strong nuclear force of attraction acts between every nucleon.
- Nuclear fission occurs when a nucleus is made to split and release a number of neutrons. This can be induced by striking a fissile nucleus with a neutron. A relatively large amount on energy is released during this process.
- When fission occurs, the mass of the fission fragments is always less than the mass of the original particles. This decrease in mass is proportional to energy, as given by $E = mc^2$.
- The energy of subatomic particles is measured in electronvolts (eV).
- 1 eV = 1.6×10^{-19} J
    - To convert from eV to joules: multiply by 1.6×10^{-19}.
    - To convert from joules to eV: divide by 1.6×10^{-19}.

KEY QUESTIONS

1 What is the strong nuclear force?

2 Why is dangerous waste produced by nuclear reactors?

3 Consider one particular neutron in the nucleus of a gold atom. Describe the forces that the neutron experiences from other nucleons.

4 Convert 5.0 MeV into joules.

5 Convert 6.0×10^{-15} J into eV.

6 Which of these nuclides below are fissile and which are non-fissile?
cobalt-60, uranium-235, uranium-238, plutonium-239

7 Determine the number of neutrons (x) released in this fission reaction:

$${}^{1}_{0}n + {}^{235}_{92}U \rightarrow {}^{148}_{57}La + {}^{85}_{35}Br + x\,{}^{1}_{0}n$$

8 The mass defect during the process in question 7 is 2.12×10^{-28} kg. Give your answers to three significant figures.

- **a** Calculate the energy (in joules) released in this fission process.
- **b** Express this energy in electronvolts.

7.2 Nuclear fusion

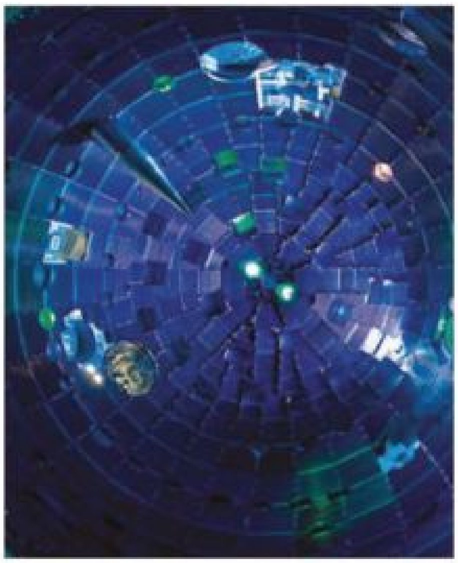


FIGURE 7.2.1 The experimental fusion reactor at the National Ignition Facility in the USA.

Nuclear fusion is a process that has been occurring inside the Sun and other stars for billions of years. **Fusion** involves the combining of small nuclei such as hydrogen and helium to form a larger nucleus. The amount of energy released per nucleon with fusion is greater than with fission and there is no radioactive waste produced. Scientists are working on experimental fusion reactors such as ITER (International Thermonuclear Experimental Reactor) in France and the National Ignition Facility in the USA, which is shown in Figure 7.2.1.

The fusion process created so far has only lasted for a fraction of a second and it is not expected that a fusion reactor will successfully operate for many years. Researchers at Lockheed Martin in the USA are working on a compact fusion reactor. In 2014 they claimed a prototype will be running by 2019. This claim has been met with scepticism by some in the scientific community.

In the study of nuclear reactions, Einstein's idea of mass–energy equivalence is used to explain why energy is released during the splitting of a nucleus. When fission occurs, the mass of the particles in the fission fragments is lower than that of the parent nucleus. Einstein's famous equation, $E = mc^2$, can be used to calculate the amount of energy that is related to this missing mass. This equation also applies to fusion reactions. It is important to note that less than 1% of matter is converted into energy in all of the energy transformations being discussed here.

NUCLEAR FUSION

Nuclear fusion occurs when two light nuclei are combined to form a larger nucleus. The example of nuclei fusing to form a helium atom is shown in Figure 7.2.2(a).

As in the cases of radioactive decay and nuclear fission, the mass of the reactants is slightly greater than the mass of the products when the nuclei combine during fusion. This mass difference is represented by the unbalanced scales shown in Figure 7.2.2(b).

The energy created by this missing mass can again be determined from:

$E = mc^2$

where E = energy (J)

m = mass defect (kg)

$c = 3.0 \times 10^8$ m s^{-1}

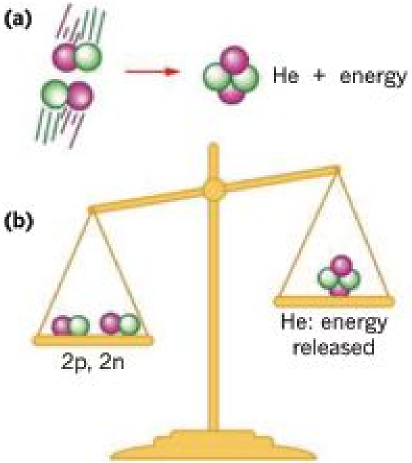


FIGURE 7.2.2 (a) When two isotopes of hydrogen fuse to form a helium nucleus, energy is released. (b) The loss in mass, m, can be calculated using $E = mc^2$.

Nuclear fusion is a very difficult process to recreate. The main problem is that nuclei are positively charged. They exert an electrostatic force of repulsion on each other, that is, they push each other away. As such, it is not easy to force the nuclei together. Remember that the electrostatic force is a long-range force and the strong nuclear force of attraction only acts at much shorter distances.

As two nuclei approach each other, the electrostatic force will cause them to be repelled. Slow-moving nuclei with relatively small amounts of kinetic energy will not be able to get close enough for the strong nuclear force to come into effect. Fusion will not happen, as can be seen in Figure 7.2.3(a).

If the nuclei are travelling towards each other at higher speeds, as shown in Figure 7.2.3(b), they may have enough kinetic energy to overcome the repulsive force. The nuclei can now get close enough for the strong nuclear force to start acting. If this happens, fusion will occur.

The graph in Figure 7.2.4 shows the effect of the electrostatic force and the strong nuclear force on the potential energy of a pair of deuterium (^2_1H) nuclei. At large separation distances, the electrostatic force dominates and the nuclei repel each other (shown to the right of the energy barrier in the graph). At small separation distances, the strong nuclear force dominates and the nuclei can fuse together. However, to get the nuclei to this point, they need an enormous amount of energy. Temperatures in the hundreds of millions of degrees are required. This enormous amount of energy enables the nuclei to overcome the energy barrier shown in the graph and fuse together.

As in fission, in any fusion reaction the atomic numbers and mass numbers on either side of the equation are conserved. The fusion of two hydrogen-2 (deuterium) nuclei is shown below:

$$^2_1H + ^2_1H \rightarrow ^3_1H + ^1_1H$$

The atomic numbers add up to two on both sides and the mass numbers add up to 4 on both sides. However, the total mass of the reactants will be greater than the total mass of the products. The mass defect is converted into energy according to $E = mc^2$.

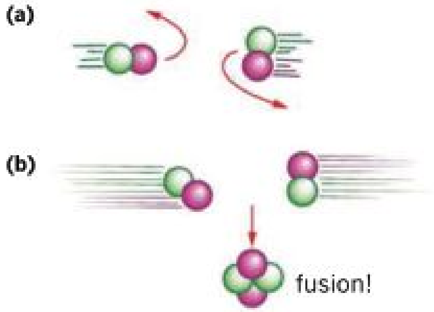


FIGURE 7.2.3 (a) Slow-moving nuclei do not have enough energy to fuse together. The electrostatic forces cause them to be repelled from each other. (b) If the nuclei have sufficient kinetic energy, then they will overcome the repulsive forces and move close enough together for the strong nuclear force to come into effect. At this point, fusion will occur and energy will be released.

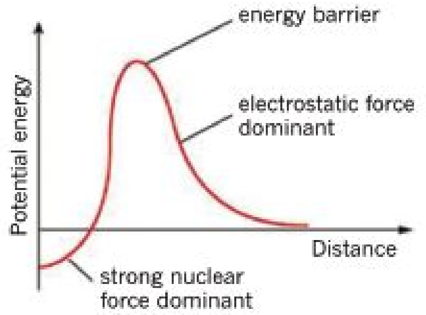


FIGURE 7.2.4 If two hydrogen-2 (deuterium) nuclei are to get close enough for the strong nuclear force to act, they must overcome the energy barrier presented by the electrostatic force.

PHYSICSFILE

Hydrogen bomb

In 1952, a fusion reaction was used to power the world's first hydrogen bomb. It had five times the destructive power of all the conventional bombs that were dropped during the whole of World War II. The high temperature achieved by a fissile fuel explosion was used to initiate the fusion reaction. In other words, an atomic bomb was used as the fuse for the hydrogen fusion bomb.

FIGURE 7.2.5 The hydrogen bomb dropped at Bikini Atoll in 1956.

BINDING ENERGY

The total mass of a stable nucleus is slightly less than the combined mass of the individual protons and neutrons. Einstein realised that this mass defect is linked to energy using: $E = mc^2$.

- During nuclear fusion when a nucleus forms, mass is lost and converted into energy.
- During nuclear fission when an unstable nucleus splits into fragments, mass is lost and converted into energy.
- The energy released per nucleon is greater during fusion than during fission.

The mass defect in nuclei can be converted into energy using $E = mc^2$. This is known as the binding energy of the nucleus. The binding energy of the nucleus indicates how much energy is needed to separate the nucleus into individual protons and neutrons.

Each nucleus has its own binding energy value, with a higher value indicating a more stable nucleus. A binding-energy-per-nucleon graph, as shown in Figure 7.2.6, allows a comparison of nuclear stabilities.

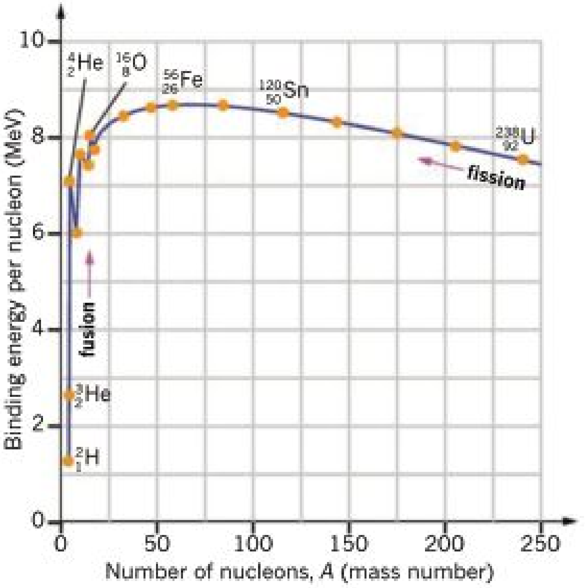


FIGURE 7.2.6 The graph of binding energy per nucleon shows that elements with mass numbers between 40 and 80 are the most stable.

Figure 7.2.6 can be analysed to better understand fission and fusion.

- Small nuclei have lower binding energy per nucleon values, indicating that they are easier to break apart compared to larger nuclei. Helium-4 has a relatively high value indicating that it is stable.
- The binding energy per nucleon increases dramatically for the very small nuclei. As they fuse together, the binding energy per nucleon increases. This is the energy released during fusion.
- Elements with mass numbers between 40 and 80 have nuclei that are tightly bound. It takes more energy to break these nuclei apart. These are the most stable nuclei. These elements have the highest binding energy per nucleon values on the graph.
- Larger nuclei have lower binding energy per nucleon values, indicating relatively lower stabilities.
- If a large nucleus such as uranium splits into two fragments, the binding energy per nucleon of the fragments again increases. This is the energy released during fission.
- Iron (Fe), with a mass number of 56, has the most stable nucleus.
- Nuclei smaller than iron undergo fusion and release energy. Nuclei larger than iron undergo fission and release energy.

7.2 Review

SUMMARY

- Nuclear fusion is the combining of light nuclei to form heavier nuclei. Extremely high temperatures are required for fusion to occur.
- A small amount of mass is lost during a fusion reaction. This mass is related to the energy produced according to $E = mc^2$.
- Nuclei are positively charged and so repel each other due to electrostatic forces. Approaching nuclei must have enough speed to overcome the electrostatic forces and get close enough for the strong nuclear force to take effect. The energy that is required to achieve this is called the energy barrier.
- The amount of energy released per nucleon is greater for fusion than for fission processes.
- Binding energy indicates how much energy is needed to separate the nucleus into individual protons and neutrons.
- The binding energy per nucleon increases dramatically for the very small nuclei. As nucleons fuse together, the binding energy per nucleon increases. This is the energy released during fusion.
- If a large nucleus such as uranium splits into two fragments, the binding energy per nucleon again increases. This is the energy released during fission.

KEY QUESTIONS

1 How does fusion differ from fission?

2 The fusion reaction that is most promising for use in nuclear fusion reactors is:

$${}^2_1\text{H} + {}^3_1\text{H} \rightarrow {}^4_2\text{He} + {}^1_0\text{n}$$

Why is energy released during this process?

3 How does the amount of energy released per nucleon during a single nuclear fission reaction compare to the amount of energy released per nucleon for a single fusion reaction?

4 During the process of nuclear fusion, mass is lost and this appears as energy. What is the approximate percentage of mass lost during a typical fusion reaction?

5 Consider the fusion reaction shown below.

$${}^2_1\text{H} + {}^3_1\text{H} \rightarrow {}^a_b\text{X} + {}^1_0\text{n}$$

- **a** Determine the values of *a* and *b* and hence the symbol of the unknown element X.
- **b** 33 MeV of energy is released. What is the mass defect in kg? Give your answer to two significant figures and in scientific notation.

6 Two slow-moving protons are travelling directly towards each other. Explain whether the protons will collide and fuse together.

7 Two fast-moving protons are travelling directly towards each other. The protons collide and fuse together. Explain why this happens.

8 The following fusion reaction is taking place in the Sun.

$${}^3_2\text{He} + \text{X} \rightarrow {}^4_2\text{He} + {}^1_1\text{p} + {}^1_1\text{p}$$

During each fusion reaction, 23 MeV of energy is released.

- **a** What are the atomic and mass numbers of particle X and what is its symbol?
- **b** Convert 23 MeV into joules. Give your answer to two significant figures and in scientific notation.
- **c** Determine the mass defect for this fusion process. Give your answer to two significant figures and in scientific notation.

9 What happens to the binding energy and the stability of two hydrogen-2 nuclei when they are fused together to form helium-4?

10 What happens to the number of nucleons during the fusion reaction below?

$${}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + {}^1_0\text{p}$$

7.3 Electromagnetic waves and synchrotron radiation

Radiation can be released from the nuclei of atoms during the process of radioactive decay. One form of radiation discussed previously is gamma rays. In this section, another fundamental type of radiation—light—will be examined. Gamma rays and visible light are both forms of electromagnetic radiation. Visible light is released from atoms, but not from the nuclei.

The very high energy electromagnetic radiation that is produced by a synchrotron will also be covered in this section. Australia's only synchrotron is located near Monash University in Clayton, Melbourne, and is pictured in Figure 7.3.1.

FIGURE 7.3.1 The Australian Synchrotron is about the size of a football ground. This size is necessary to contain the electrons which are travelling at almost the speed of light as they zoom around the storage ring.

ELECTROMAGNETIC WAVES

In the late 1600s, light was known to involve the transfer of energy from one place to another. In Isaac Newton's time, a particle model and a wave model for light had both seemed equally valid. In the early 1800s, as the result of considerable efforts by scientists, one model prevailed. The wave model for light was universally accepted. At this time, the speed of light, c, could also be measured for the first time, at 300 000 km s^{-1}.

PHYSICSFILE

The photon model

The accepted modern model for light involves the properties of both waves and particles. This model incorporates the concept of a wave-like particle called the **photon**. This model is explored in detail in Unit 4.

By the 1860s investigations being carried out on different forms of **electromagnetic radiation** (EMR) led to the finding that visible light itself is just one of the many forms of EMR. The electromagnetic spectrum includes radiation at a wide range of frequencies (or wavelengths).

All electromagnetic waves are created by accelerating charges. Charges, even when they are not moving, are associated with an electric field. When the charges are made to accelerate, the result is two fields. These fields are rapidly changing magnetic and electric fields travelling out from the source at the speed of light. The electric field component and the magnetic field component are at right angles to each other and to their direction of travel, as shown in Figure 7.3.2.

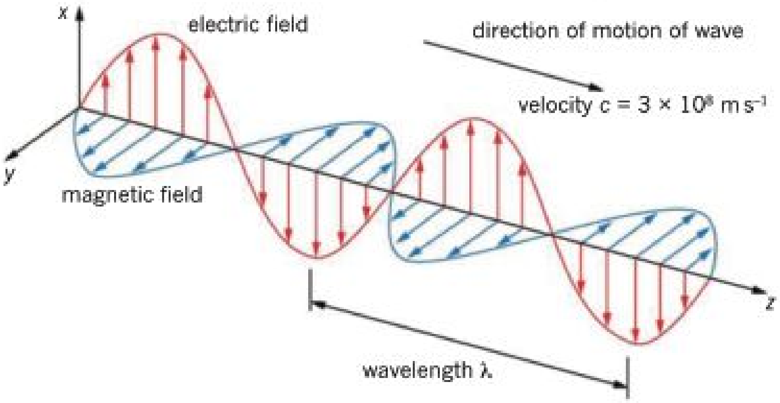


FIGURE 7.3.2 As all electromagnetic waves travel with the same speed, the only thing that differentiates one form of EMR from another is the frequency and, therefore, the wavelength.

The electromagnetic spectrum

The many forms of EMR are essentially the same, differing only in their frequency and wavelength. The electromagnetic spectrum is divided into categories depending on how the radiation is used and its frequency. These categories are shown in Figure 7.3.3.

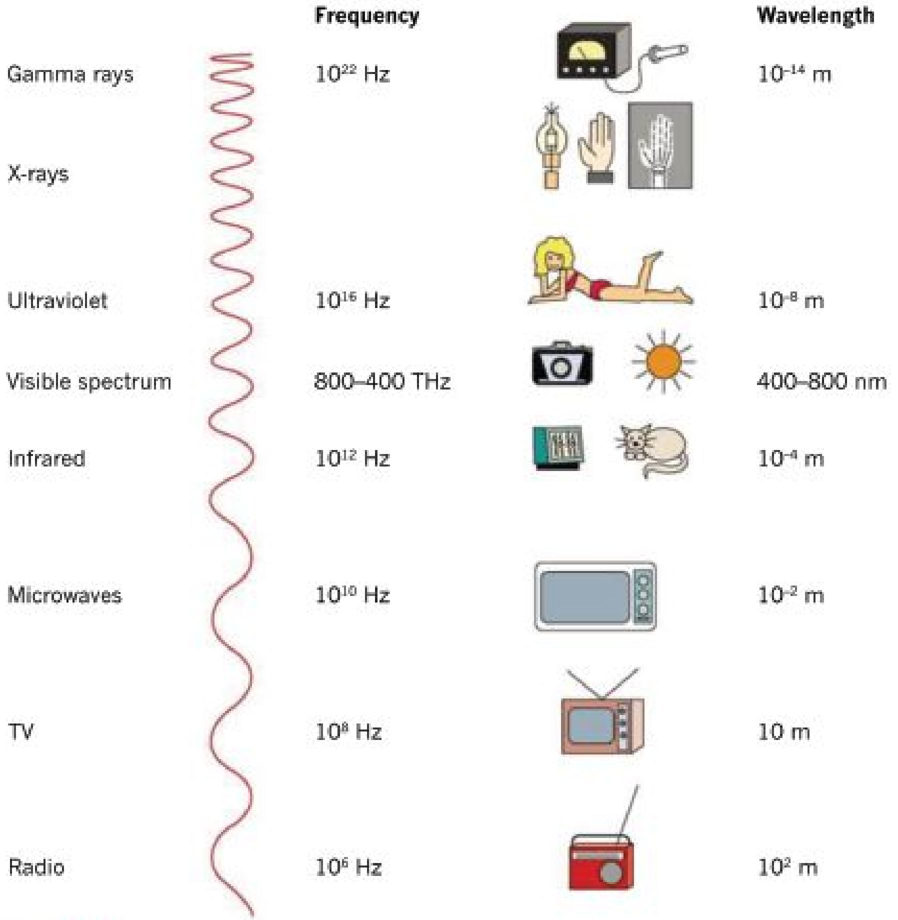


FIGURE 7.3.3 The electromagnetic spectrum. All of these types of radiation travel at the speed of light.

The energy carried by the electromagnetic radiation is proportional to its frequency. High-frequency, short-wavelength gamma rays are at the high-energy end of the spectrum. Low-frequency, long-wavelength radio waves carry the least energy.

Humans have cells in their eyes that can respond to EMR of frequencies between approximately 400 THz and 800 THz (1 terahertz or 1 THz = 10^{12} Hz). These frequencies make up the visible light section of the electromagnetic spectrum. Visible light is the region with wavelengths from 400 to 800 nm (1 nanometre or 1 nm = 10^{-9} m). The human eye cannot perceive any wavelengths of EMR outside this range, although some other animals are capable of seeing ultraviolet and infrared radiation.

Visible light can be produced by the rearrangement of electrons in atoms and molecules. This may be caused, for example, by heating a tungsten filament in an old light globe like that in Figure 7.3.4. Visible light can also be produced when electrons rearrange in LEDs and fluorescent tubes.

FIGURE 7.3.4 This tungsten light globe filament becomes hot as electricity passes through it. The charges in the hot metal accelerate, producing electromagnetic radiation. The globe produces about 90% infrared radiation and about 10% visible light energy. This is a very inefficient way of producing light and filament globes have largely been replaced with compact fluorescent tubes and LED globes.

Other forms of electromagnetic radiation are produced as described below.

- Radio waves are produced by charges accelerating through wires. Large radio-tower antennae are used by radio stations and phone companies to transmit their signals.
- Microwaves are produced by accelerating electrons along short metal antennae. These waves are used for communications and for cooking food.
- Infrared radiation (IR) is produced by any object with thermal energy as the molecules and atoms vibrate. The Sun produces a great deal of infrared radiation, but any object above absolute zero is emitting some IR.
- Ultraviolet radiation (UV) is produced by the Sun by the movement of charged particles. It can also be created in glass tubes as electrons rearrange. Some types of UV radiation can cause skin cancer.
- X-rays are produced when fast-moving electrons are fired into a metal target and decelerate rapidly.
- As was discussed in the radioactivity section, gamma rays are released from the nucleus of some radioactive atoms.

SYNCHROTRONS

A **synchrotron** is a very large machine that produces and accelerates electrons to close to the speed of light, to be used for different research purposes. Australia's only synchrotron opened in 2007 and is located in Clayton, Victoria.

PHYSICSFILE

Synchrotrons and solving crimes

Forensic science has been identified as one of the priorities for the use of the Australian Synchrotron during its first few years of operation. The intensity and collimation (narrowness) of synchrotron light make the synchrotron an invaluable tool in successfully analysing microparticles of forensic evidence, such as skin, hair and paint, left at the scene of a crime. The FBI has employed such techniques using the beamlines of the Brookhaven Synchrotron in the USA.

A synchrotron works as follows and is shown in Figure 7.3.5.

- A synchrotron is a particle accelerator designed to circulate electrons around a closed path at speeds very close to the speed of light.
- Electrons are emitted from an electron gun in pulses. They are accelerated in bunches through the **linear accelerator** (linac) by powerful bursts of radio-frequency (RF) radiation.
- Electrons then travel around a booster ring, being accelerated further as they pass through RF cavities until reaching an energy of 3 GeV (3×10^9 eV).
- Electrons are then channelled into the storage ring. Synchrotron radiation is produced as the electrons are accelerated through curved paths steered by the strong magnetic fields of the dipole magnets and insertion devices.
- A conical beam of synchrotron radiation travels at a tangent to these accelerating electrons and passes down one of the **beamlines** to one of the experimental stations.

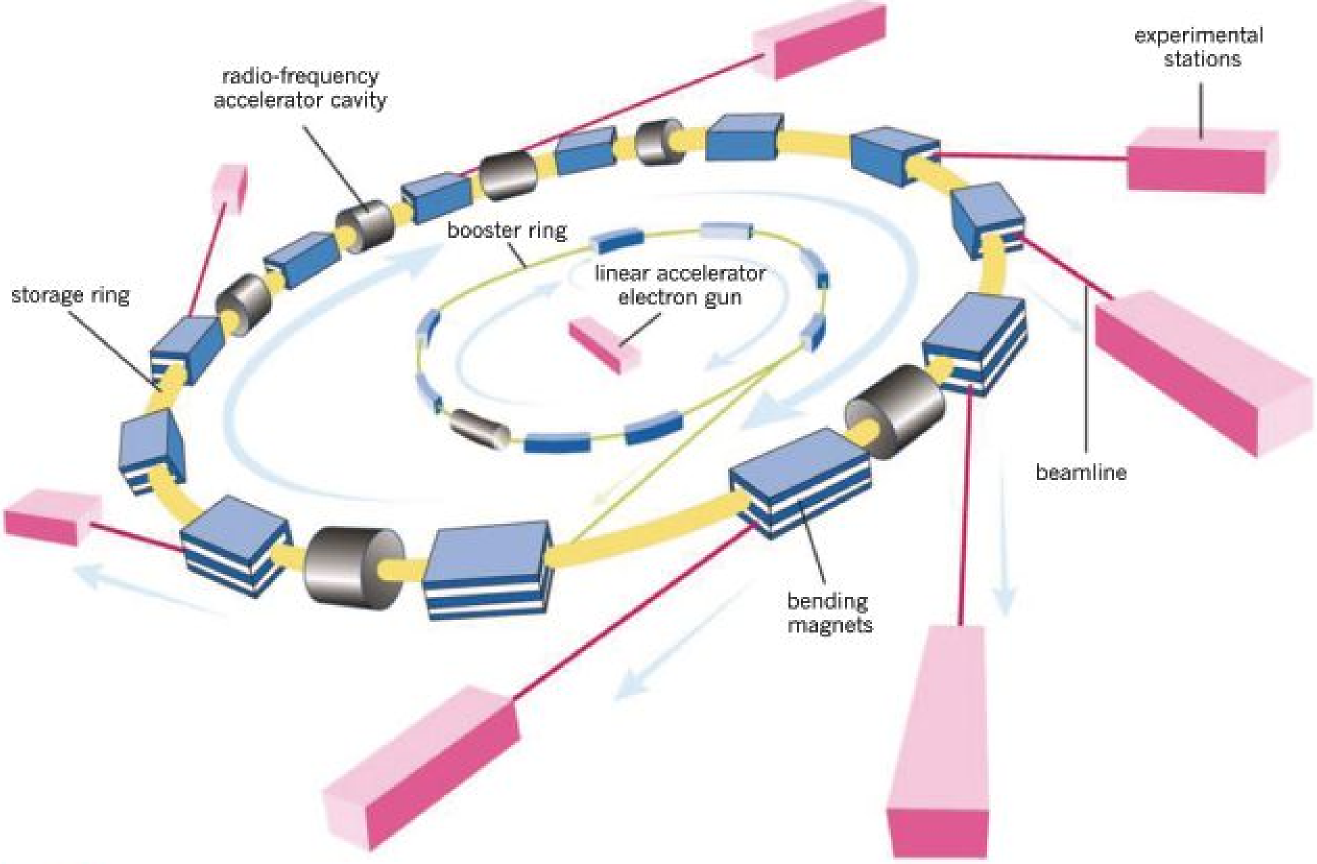


FIGURE 7.3.5 A model of the inner workings of a synchrotron, like the Australian Synchrotron in Melbourne.

Synchrotron light

Synchrotron light is the name given to electromagnetic radiation emitted when charged particles, such as electrons, are accelerated in curved paths in a synchrotron machine. For high-energy electrons, the electromagnetic radiation emitted has energies ranging from the infrared through to X-rays. Synchrotron-generated light has a number of advantages that make it suitable for a range of experimental techniques. It has:

- a broad spectral range, from microwaves to X-rays
- a high intensity or brightness
- a well-defined, coherent (synchronised) beam
- the potential to be tuned to required frequencies.

PHYSICSFILE

Medical applications

There is an enormous amount of genetic information that can be extracted from the DNA analysis of a single strand of hair. Now, there are more secrets your hair could reveal. An Australian researcher has observed that hair from women at a higher risk of developing breast cancer produces a different X-ray diffraction pattern compared with that of other women's hair. It is an exciting possibility for the future that a screening test could be as simple as examining a single hair with synchrotron light.

The range of energies produced in the Australian Synchrotron can be seen in Figure 7.3.6.

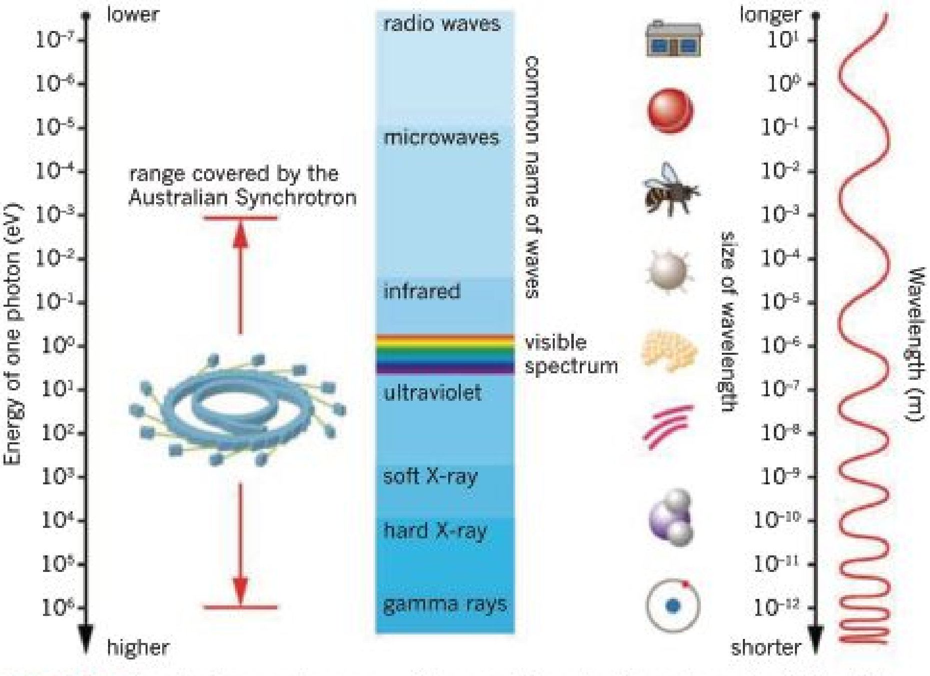


FIGURE 7.3.6 Synchrotrons produce a very wide range of wavelengths and energies. At the high energy end, the wavelength of the synchrotron light is as small as the sizes of atoms and molecules.

The specific attributes of synchrotron radiation make it useful in a wide range of research techniques. The electromagnetic radiation it produces has wavelengths corresponding to the dimensions of cells, viruses, proteins and atoms.

Shorter wavelengths enable scientists to explore the structure of similarly small objects. An ordinary optical microscope is incapable of resolving such structures, due to the much longer wavelength of visible light. (Visible light has a wavelength much longer than the size of proteins.)

The short-wavelength X-rays produced in synchrotron light are an ideal tool for examining structures at a cellular or atomic level. High-energy X-rays can resolve down to scales of 10^{-10} m, the size of individual atoms.

PHYSICS IN ACTION

Lithography

Everyday appliances are getting smaller. The ability to make extremely small devices could lead to groundbreaking new applications in medicine, microsensors and micromachining.

Lithography consists of printing a pattern onto another surface. Lithographic techniques have existed for centuries. The method of using X-rays to etch a pattern is much more recent. X-ray lithography is an exciting application of synchrotron radiation. Lithographic techniques can be used to create tiny mechanical parts or to produce moulds that may be filled with a range of materials, including metals, glass or ceramics, to make components for devices. This technique is capable of producing high-resolution images between 100 μm and 2 mm thick.

One application is the production of micro-electrical mechanical devices such as microsensors and actuators, for example, motion devices, pumps and filters. Synchrotron light is suitable for providing the X-rays used in the lithographic process.

X-rays of wavelength 0.01–1.0 nm are generally selected from the beam. The X-rays are shone through a patterned mask onto a photosensitive coating (called a photoresist) that is later developed to reveal the image. This process is shown in Figure 7.3.7. As the radiation is exposed, it casts a shadow of the mask pattern onto the resist. X-rays interact with the resist coating. This photosensitive layer produces images of resolution currently approaching 0.14 μm. Electrons produced from the photoelectric effect induce a change so that the exposed resist can be removed by chemical development. The unexposed resist is left behind. Imaging systems called X-ray aligners, or steppers, are used to produce a clearer image. They operate on a vertically held mechanical stage.

The coherent X-ray beam produced in synchrotron radiation makes it particularly suited to the technique of X-ray lithography. The shadow-blurring effect due to light incoherence is reduced. Because the beam is of high intensity, shorter exposure times are used. The suitability of synchrotron light to this process has driven the fabrication of much smaller, more compact storage rings specifically to perform X-ray lithography work. A process of deep-etch lithography, called LIGA, can be used to produce gears only a few thousandths of a millimetre in size (see Figure 7.3.8). These components, along with microscopic turbines, can be used to build microelectronic devices in the revolutionary new field of micromachining.

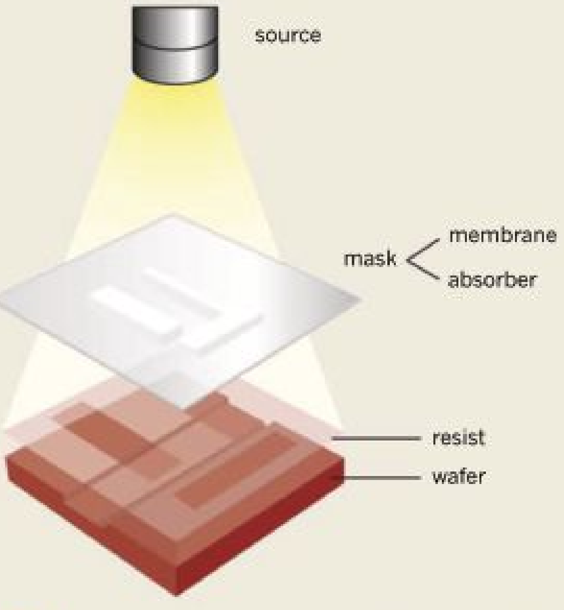


FIGURE 7.3.7 The basic set-up of an X-ray lithographic process. The mask used in the process consists of a transparent part called the carrier. This is a thin membrane, generally 1–2 μm thick, made from silicon carbide or silicon nitride. The pattern to be copied is structured in a thick layer of a heavy element that is attached to this membrane. This section is called the absorber, and is usually made up of gold or tungsten. X-rays are absorbed into the absorber and pass through the membrane in order to transmit the desired pattern to the wafer below.

FIGURE 7.3.8 Gears like this one can be produced by using deep-etch lithographic techniques. They are used to build micromachines.

7.3 Review

SUMMARY

- Electromagnetic waves are created by accelerating charges. The accelerating charges create a rapidly fluctuating magnetic field and electric field travelling out from the source at the speed of light, c (3.00×10^8 m s^{-1} in a vacuum).
- The electromagnetic spectrum consists of radio waves, microwaves, infrared, visible light, ultraviolet, X-rays and gamma rays.
- Visible light is the region with wavelengths from 400 to 800 nanometres.
- A synchrotron is a particle accelerator designed to circulate electrons around a closed path at speeds very close to that of light.
- Synchrotron radiation is produced as the electrons are accelerated through curved paths steered by strong magnetic fields.
- Synchrotrons can produce radiation with a broad spectral range—from microwaves to X-rays.

KEY QUESTIONS

1 What is electromagnetic radiation?

2 Describe what happens to the wavelength of electromagnetic radiation as its frequency increases.

3 To what is the energy carried by electromagnetic radiation proportional to?

4 What is the approximate wavelength range of visible light?

5 At the Australian Synchrotron, high-energy electrons travel in a curved path under the influence of a magnetic field. What is the spectral range of the synchrotron radiation?

6 How is synchrotron radiation created?

7 How do the speeds of radio waves, microwaves and X-rays compare?

8 A photon of light from a synchrotron has 2.5×10^{-14} joules of energy. Convert this value into electronvolts.

9 What are the main functions for the following components of the Australian Synchrotron?

- **a** linac
- **b** booster ring
- **c** storage ring
- **d** beamline

10 Compare the speeds of the electron beam in a synchrotron and the synchrotron radiation that it produces.

7.4 The production of light

Visible light is only a very small part of the electromagnetic spectrum. The human eye has evolved to be sensitive to this part of the electromagnetic spectrum. Visible light can be produced as a result of extreme heat. The flame of a candle produces infrared radiation and visible light as a result of the thermal energy causing charged particles in the flame to accelerate.

In houses today, compact fluorescent tubes and LEDs are commonly used as light sources. LEDs are increasingly being used in traffic lights, car lights and in street lights like those in Figure 7.4.1. These use a completely different and more efficient process to produce light compared with candles. The light produced by LEDs is created by interactions involving the electrons surrounding the nucleus. If you have ever used a laser pointer or scanned a barcode at a checkout, you have made use of this process.

FIGURE 7.4.1 Sodium vapour lamps have been used for street lighting for many years. The characteristic orange colour is the result of sodium's emission spectrum. These are being replaced by LEDs which produce white light and are much more energy efficient. In both types, light is produced by electron transitions between energy levels.

EMISSION SPECTRA AND ENERGY LEVELS IN ATOMS

Towards the end of the nineteenth century, scientists had devised a variety of ways of making atoms produce light. These methods included:

- heating substances until they glowed
- applying high voltages to gases in glass tubes causing the gas to glow
- burning salts in gas flames causing the salt to produce a bright flash of light.

When the light that was emitted from the atoms was analysed using a spectroscope, a distinctive **emission spectrum** was observed for each different atom. Thus the emission spectrum of an atom became a unique property of the atom. See, for example, the two emission spectra in Figure 7.4.2, which can be used to distinguish between sodium and mercury.

Recall that white light is actually made up of an infinite number of different frequencies (or wavelengths) of light. If white light from an incandescent light globe is passed through a spectroscope, a continuous rainbow of colours is seen. The rainbow will contain all the shades of the visible light spectrum from red to violet. The emission spectra for sodium and mercury in Figure 7.4.2 do not show continuous rainbows, just some specific colours (frequencies). For many years, scientists could not explain why atoms emitted only discrete frequencies of light rather than continuous spectra.

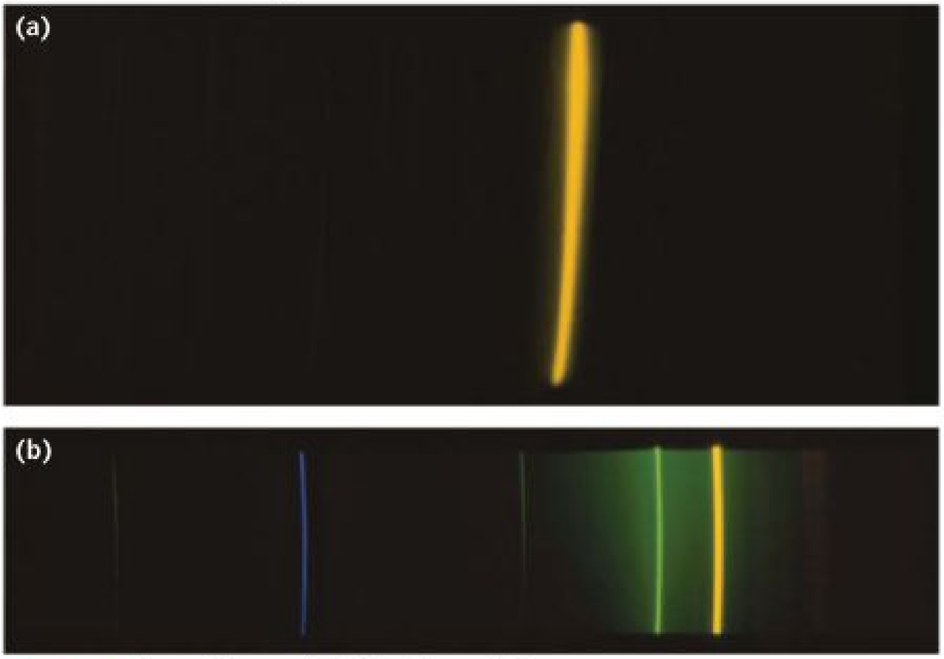


FIGURE 7.4.2 The emission spectra of (a) sodium and (b) mercury.

In 1912, Niels Bohr devised a sophisticated model of electron **energy levels** for the atom. He was later awarded a Nobel Prize in Physics for this work. Energy levels can be shown as a number of horizontal lines on a graph. The graph in Figure 7.4.3 shows the energy levels for sodium gas.

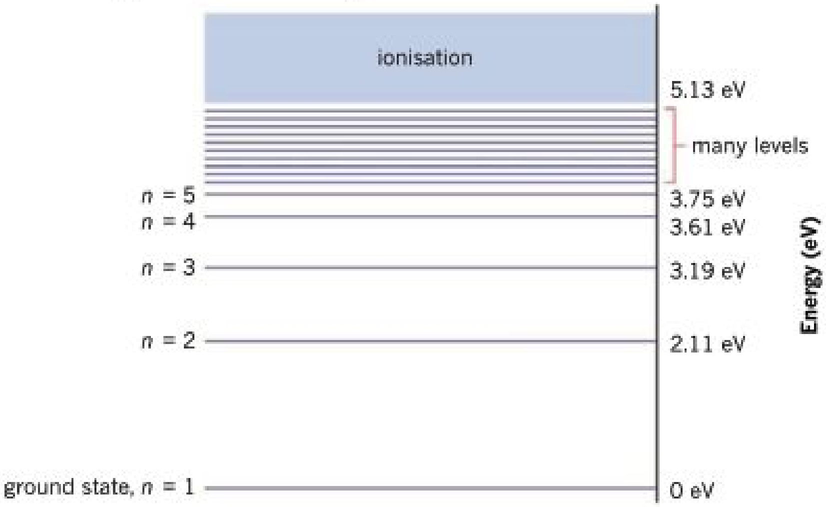


FIGURE 7.4.3 Energy levels in a sodium atom. $n = 1$ is the ground state, and the highest level is where ionisation occurs and the electron escapes the atom.

Bohr's main ideas were as follows.

- The electron moves in a circular orbit around the nucleus of an atom.
- The force keeping the electron moving in a circle is the electrostatic force of attraction (the positive nucleus attracts the negative electron).
- A number of allowable orbits of different radii exist for each atom and are labelled $n = 1, 2, 3$, etc. The electron may occupy only these orbits.
- An electron ordinarily occupies the lowest energy orbit available.
- An electron can jump to a higher energy level by absorbing some energy. The absorbed energy must be exactly equal to the difference between the electron's initial and final energy levels.
- Electromagnetic radiation is emitted by an excited atom when an electron falls from a higher energy level to a lower energy level. The energy of the emitted light will be exactly equal to the energy difference between the electron's initial and final levels.

These ideas are shown in Figure 7.4.4. In this particular example, the electron absorbs energy from light that strikes the atom. The energy absorbed is just the right amount for the electron to make the jump from its ground state to a higher level. In this diagram, the light is labelled as a photon.

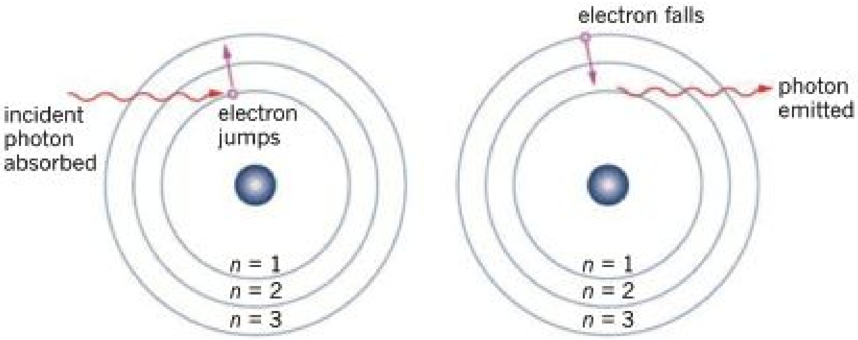


FIGURE 7.4.4 (a) If the incident photon (light) carries an amount of energy equal to the energy difference between two levels, the photon's energy can be absorbed, allowing the electron to jump to the higher level. The photon ceases to exist. (b) An atom will remain in an excited state for less than a millionth of a second. The electron will then fall to its ground state. The electron may fall in one step, or in a number of stages, emitting a photon or photons as it falls.

Each possible electron transition (jump) produces light of different energy. The energy corresponds to a different coloured line in the emission spectrum for that atom. The greater the energy emitted, the higher the frequency of the light.

PHYSICSFILE

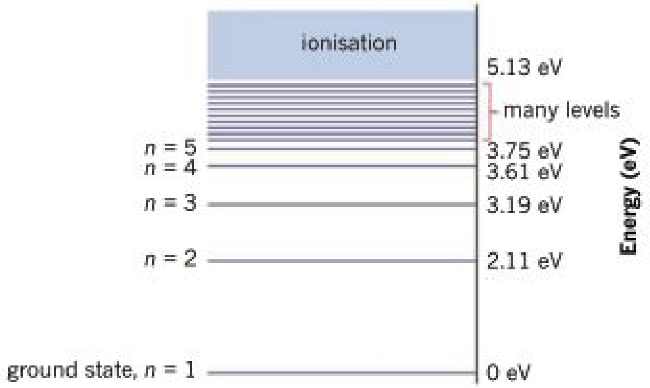
Labelling energy levels

There are two systems in use for labelling the energy levels of an atom. Sometimes the ground state ($n = 1$) is allocated 0 eV and therefore the higher levels have positive values. Alternatively, sometimes the ground state is allocated a negative value of eV and the ionisation energy level ($n = \infty$) has a value of 0 eV.

Worked example 7.4.1

ENERGY LEVELS

The energy levels for sodium gas are shown. Work out the energy of the light that is produced as an electron drops from the $n = 4$ to the $n = 2$ state.

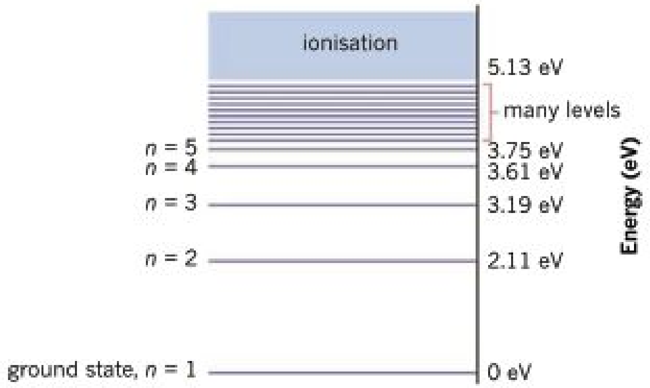


Thinking	Working
Using the figure, find the energy (in eV) of each level involved.	$n = 4$, $E_4 = 3.61$ eV $= 2$, $E_2 = 2.11$ eV
Calculate the difference between these levels.	$\Delta E = E_4 - E_2$ $= 3.61 - 2.11$ $= 1.50$ eV

Worked example: Try yourself 7.4.1

ENERGY LEVELS

The energy levels for sodium gas are shown. Work out the energy of the light that is produced as an electron drops from the $n = 4$ to the $n = 3$ state.



EXTENSION

How lasers work

Lasers use the principle of stimulated emission of radiation. In fact, laser is an acronym for Light Amplification by Stimulated Emission of Radiation. Lasers produce a coherent and very intense light beam that is the result of a chain reaction, where photons cause identical photons to be produced.

There are a number of conditions necessary for laser light to be produced. First, there must be more atoms in the excited state than in the ground state. This is not the normal condition of matter and it is achieved by 'pumping' the lasing material with energy from an external source. The energy to excite the atoms is provided by the bombarding electrons (electrical pumping) or the bombarding photons (optical pumping).

The second required condition is that the excited state used must be metastable. This means that the excited state must be a relatively stable energy level where the electron takes longer than normal to de-excite and drop to a lower level.

Finally, the photons emitted as the electrons jump back down to ground state must be used to continue the chain reaction of photon emissions. This is achieved by placing mirrors at the end of the laser tube. One mirror reflects all the photons back, and the other mirror is partially silvered, making it reflect some, but not all, of the photons. The mirrors cause the photons to be reflected back and forth through the lasing medium, stimulating further emissions. This is the 'light amplification' part of the process. At the same time, some light escapes through the partially reflective mirror. This is the laser beam. The entire process is illustrated in Figure 7.4.5.

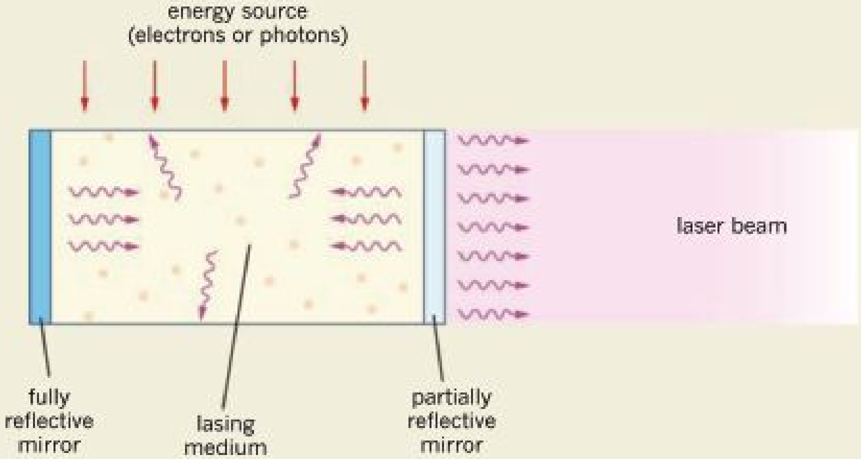


FIGURE 7.4.5 As the emitted photons reflect back and forth between the mirrors, further stimulated emissions occur and a laser beam is produced.

7.4 Review

SUMMARY

- An emission spectrum is produced by energised atoms as electrons undergo transitions between energy levels. The spectrum for an element is unique.
- Niels Bohr suggested that electrons in atoms orbit the nucleus in specially defined energy levels, and no radiation is emitted or absorbed unless the electron can jump from its energy level to another.
- An electron in an atom which drops between energy levels emits a photon (light) of energy equal to the difference between the energy levels. The energy of the photon determines the colour of the light.

KEY QUESTIONS

1 Describe a typical emission spectrum for an element.

2 What is the link between an atom's emission spectrum and its structure?

3 How are the energy levels within an atom commonly represented on a graph?

4 What is the name of the lowest energy level within an atom?

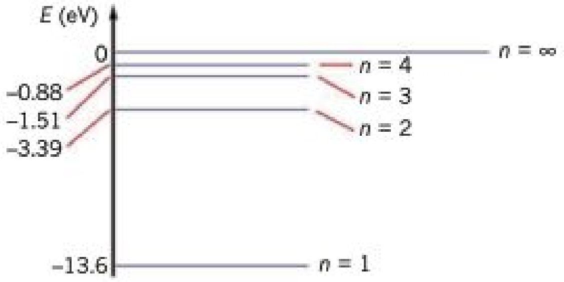
5 Can an atom emit any energy photons (light) when an electron falls from a higher energy level to a lower energy level? Explain your answer.

6 Students are using spectroscopes to analyse the light emitted by a light globe and a sodium vapour lamp. Compare their observations.

7 How is the orange light that is produced by a sodium vapour street light created?

8 Refer to Figure 7.4.3 on page 246. As electrons drop back to the ground state from $n = 3$, how many different photon energies can be produced?

Use the graph below to answer questions 9 and 10.



9 Some of the energy levels for hydrogen are shown in the graph above. Which levels are the

a ground state?

b ionisation level?

c excited states?

10 An electron is in the $n = 2$ excited state. What is the energy of the photon that is released as the electron drops back to the ground state?

Chapter review

07

KEY TERMS

beamline
binding energy
electronvolt
emission spectrum
energy levels
fissile
fission
fission fragments
fusion
linear accelerator
photon
synchrotron
synchrotron light
transuranic

1 Give the meaning of the term 'fissile'.

2 Are all atoms fissile? Give examples to support your answer.

3 Consider one particular proton in the nucleus of a zinc atom. Describe the forces that the proton experiences from other nucleons.

4 Both neutrons and α-particles can be used to trigger nuclear fission. Explain why neutrons are better than α-particles for inducing fission.

5 Einstein said that mass and energy are equivalent. In one particular fission reaction, a decrease in mass of 3.48×10^{-28} kg occurred. Give your answers to two significant figures.

a Express the energy equivalent of this mass in joules.
b Express the energy equivalent of this mass in MeV.

6 Determine the value of the unknown mass number x and atomic number y in this fission reaction:
$^{1}_{0}n + ^{x}_{94}Pu \rightarrow ^{132}_{54}Xe + ^{106}_{y}Xr + 4^{1}_{0}n$

7 Determine the number of neutrons (x) released in this fission reaction:
$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{132}_{50}Sn + ^{101}_{42}Mo + x^{1}_{0}n$

8 A typical fusion reaction is: $^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{1}H + ^{1}_{1}H$
Why are high temperatures such as 100 million degrees needed for this reaction to occur?

9 Consider this fission reaction of uranium-235:
$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{140}_{55}Cs + ^{93}_{37}Rb + 2^{1}_{0}n$
During this fission reaction there is a mass defect of 4.99×10^{-28} kg. How much energy in joules is produced per reaction?

10 Consider this fusion reaction:
$^{1}_{1}H + ^{3}_{2}He \rightarrow ^{4}_{2}He + ^{0}_{+1}e + \nu$
Hydrogen and helium-3 are being fused together and a helium-4 nucleus is being created along with a positron and a neutrino. 21 MeV of energy is released.

a How does the combined mass of the hydrogen and helium-3 nucleus compare with the combined mass of the helium-4 nucleus, positron and neutrino?
b Where has the energy that was released come from?
c Convert the energy into joules.
d What is the mass defect of this fusion reaction?

11 Compare the waste and the energy per nucleon produced by fusion reactors with those of fission reactors.

12 What happens to the binding energy per nucleon and the stability of the nucleus when a uranium-238 nucleus splits apart to form two smaller nuclei?

13 The binding energy per nucleon for iron (mass number 56) is higher than for other elements. What does this mean for the stability of iron nuclei?

14 How does an old tungsten filament light globe produce light?

15 X-ray light produced by a synchrotron has energy of 5.0×10^{4} eV. What is the energy of this light in joules?

16 Which form of electromagnetic radiation is released from the nucleus of radioactive atoms?

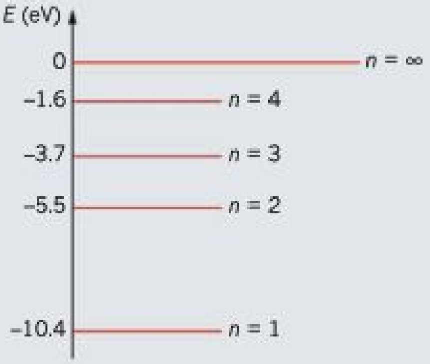
17 Which of the following is not a form of electromagnetic radiation?
UV radiation, gamma rays, beta rays, green laser light?

18 Synchrotron light is produced when electrons are accelerating around a curve in a synchrotron machine. In what direction does the light travel?

19 Refer to Figure 7.4.3 on page 246. Electrons are in the $n = 2$ excited state. As the electrons drop back to the ground state, what is the energy (in eV) of the emitted light?

20 Refer to Figure 7.4.3. Electrons in the gas are in the $n = 4$ excited state. As the electrons drop back to ground state, what is the highest energy photon of light that is produced? Answer in eV.

21 Some of the energy levels for mercury gas are shown below. Mercury atoms have been excited to the $n = 3$ level.



What are the possible energy transitions that take place when an electron at $n = 3$ returns to the ground state?

UNIT 1 • Area of Study 3

REVIEW QUESTIONS

What is matter and how is it formed?

1 A student is explaining the big bang theory to one of his friends. He tells her that at some point in time, the universe was concentrated in a very small dense mass and then, after the big bang, the mass spread out into space and continues to expand. This explanation contains several mistakes. Give a simple but more accurate explanation.

2 The same student asserts to his friend that looking at a very distant star using a telescope is effectively looking back in time. Is he correct? Explain why or why not.

3 Explain what is meant by the redshift of stellar spectra and why this is believed to be evidence that the universe is expanding.

4 The 'Coma Cluster' of galaxies is about 90 Mpc from Earth. At what speed is the cluster receding from us? Is this a significant fraction of the speed of light? (Use $H_0 = 70$ km s^{-1} Mpc^{-1}.)

5 According to the big bang theory, space itself is expanding.

a Use Hubble's law to compare the recession velocity of a star at the position of our closest star, Proxima Centauri at 1.3 pc, with that of the edge of the observable universe, which is taken to be 4.4×10^{26} m away. (Proxima Centauri is in fact moving towards the Sun because of other factors.)

b The speed of the edge of the observable universe that you have calculated exceeds the speed of light. Is that a problem? Explain.

c The Hubble constant is taken to be around 70 km s^{-1} Mpc^{-1}. Calculate the Hubble constant H_0 in SI units and explain why the redshift of a star at the distance of Proxima Centauri is not going to be measurable.

d Explain why, in contrast, stellar objects at much greater distances actually show measurable effects.

6 What is it about Cepheid variables that makes them a useful tool with which to find astronomical distances?

7 Answer the following questions about the cosmic microwave background (CMB) radiation.

a Given the extraordinarily high temperatures of the early universe, the radiation from the big bang would have been expected to be very high frequency and short wavelength gamma radiation. Explain why physicists expected the CMB radiation to be detectable in the microwave range.

b When the CMB radiation was mapped it was found that the radiation was essentially uniform in all directions, but with small variations. Explain why these small variations are considered important in stellar evolution.

8 What is 'pair production' and why is it important for understanding the creation of matter?

9 Without the phase of extremely rapid inflation of the universe, any matter created would have immediately annihilated and the universe would be empty except for the CMB radiation. Explain.

10 As the temperature of the universe dropped, it would be expected that less pair creation would occur. Explain why this would be the case.

11 Once protons, neutrons and electrons had formed, one would think that atoms would follow. Why could atoms still not exist at this point?

12 Fusion followed as protons and neutrons were forced together to form hydrogen, helium and lithium nuclei for a time. Then this process too stopped. Why?

13 What condition would the background radiation photons have to satisfy before atoms could form?

14 The universe continued to cool and expand. Explain the role of gravity in creating the first stars.

15 An electron and a positron annihilate with the production of two photons.

a Explain the source of the energy of the photons.

b Assuming the two leptons each had minimal kinetic energy, calculate the total photon energy. (Assume the mass of a lepton is 9.1×10^{-31}.)

16 Copy and complete the table about the properties of exchange particles, using the dot points below.

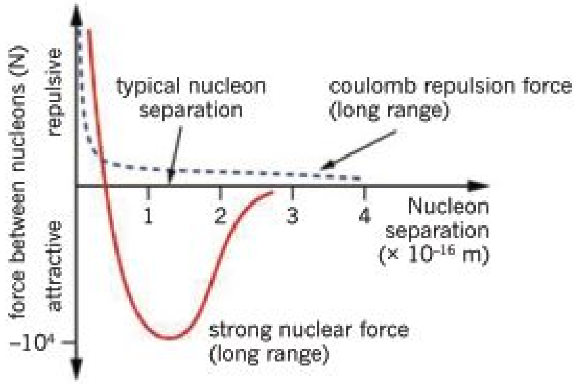
Particle	Property
gluon	
photon	
W^+, W^- and Z	
(graviton)	

- mediator of the electromagnetic force
- interacts with charged particles
- mediator of the gravitational force
- mediator of the weak nuclear force
- interacts with quarks
- causes nuclear decay
- mediator of the strong nuclear force

17 Copy and complete the table below, using the following list of particles:
electrons, positrons, neutrons, muons, (W^+, W^- and Z), neutrinos, gluons, antiprotons, photons, pions, protons and gluons

Category	Particle type	Description	Particle name
bosons		mediators of the fundamental forces	
fermions (make up all matter)	leptons	• do not experience the strong force • exchange W and Z bosons—weak nuclear force • charged leptons exchange photons—electromagnetic force	
	hadrons	• do experience the strong force, exchanging gluons • made up of quarks	
	- baryons	made of 3 quarks	
	- mesons	made of 2 quarks	

18 The figure below indicates the strong nuclear force and the electrostatic force between protons.



a Comment on the value of the typical nucleon separation in the light of the two forces above.

b Explain which particles in the nucleus each of the two forces act on.

c Explain why for increasing atomic number, the ratio of neutrons to protons increases.

19 As a result of the disaster at the Fukushima nuclear power plant in 2011, the radioactive isotopes caesium-137 and iodine-131 were released into the atmosphere. Use a periodic table to determine the number of protons, neutrons and nucleons contained in a nuclide of each radioisotope.

20 From which part of a radioisotope are the following particles emitted?

a alpha particles

b beta particles

c gamma rays

21 A radioisotope K-40 has a half-life of 1.3×10^9 years and decays to a stable isotope Ar-40, which is used for dating rocks. Copy and complete the following table.

Time ($\times 10^9$ years)	Number of K nuclei	Number of Ar nuclei	Ratio K:Ar
0	1000	0	–
1.3			
2.6			
3.9			

22 A $^{7}_{3}\text{Li}$ nucleus is bombarded with a high-speed proton resulting in the production of two identical particles. Write the nuclear equation that describes this reaction.

23 Calculate the energy (in MeV) of the following particles:

a an alpha particle with energy 1.4×10^{-12} J

b a beta particle with energy 6.7×10^{-14} J

c a gamma ray with energy 8.0×10^{-14} J.

24 A Geiger counter measures the radioactive disintegrations from a sample of a certain radioisotope. The count rate is recorded in the table.

Activity (Bq)	800	560	400	280	200	140
Time (min)	0	5.0	10	15	20	25

a Plot a graph of activity versus time.

b Use your graph to estimate the activity of the sample after 13 minutes.

c What is the half-life of this element?

d Determine the activity of the sample after 30 minutes.

25 If a particular atom in the sample has not decayed during the first half-life, which one of the following statements best describes its fate?

A It will definitely decay during the second half-life.

B It has a 50% chance of decaying during the second half-life.

C The probability that it will decay cannot be determined.

D If it does not decay during the first half-life, it will not decay at all.

26 Tritium (hydrogen-3) is radioactive and its decay equation is shown.

$^{3}_{1}\text{H} \rightarrow \text{X} + {}^{0}_{-1}\text{Y}$

a How many protons and neutrons are in each tritium nucleus?

b Which element is the daughter nuclide X?

c Which of the following best describes the nature of Y in the decay equation?

A It is a positron.

B It is an electron.

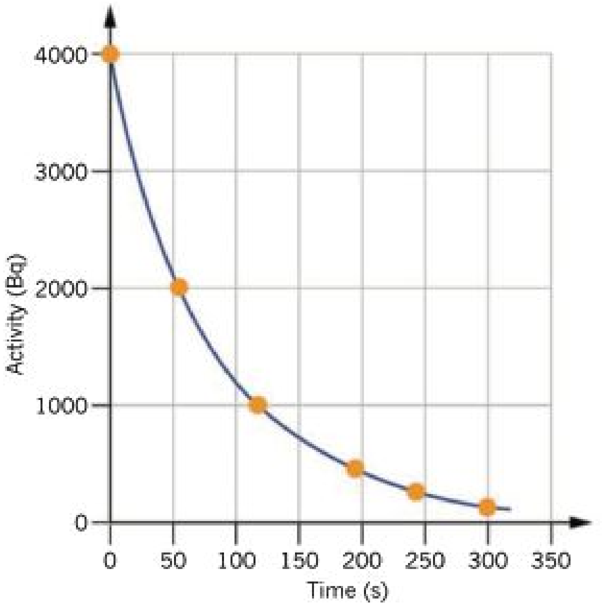
C It is a proton.

D It is a neutron.

d One of the nucleons in tritium has spontaneously transformed during this decay. Which one and what has it transformed into?

27 Given that gold is an alpha emitter, use a periodic table to help you write a decay equation for this isotope. Use $^{185}_{79}Au$

28 The graph below shows the data obtained in an experiment to determine the half-life of sodium-26.

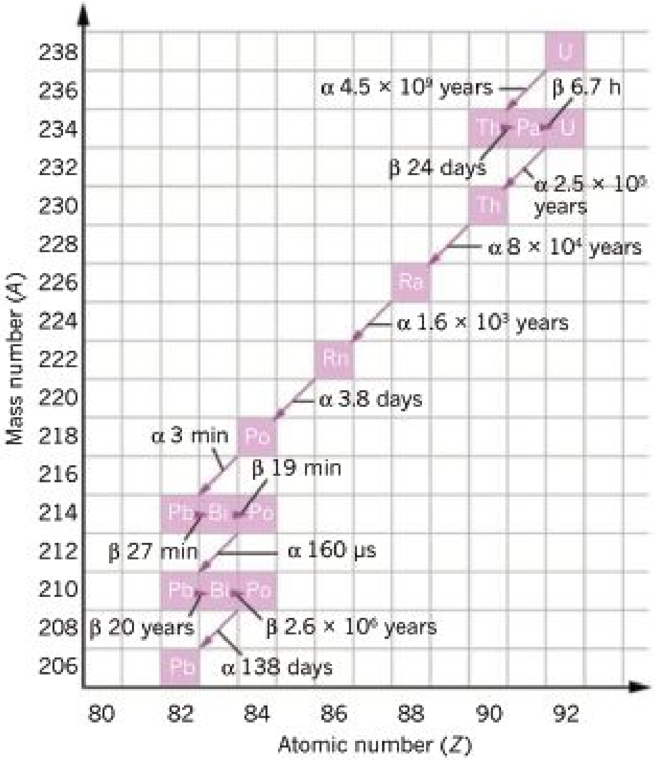


- **a** Use the graph to work out the half-life of sodium-26.
- **b** If the initial sample contained 150 g of sodium-26, how much of this radioisotope will remain after 5 minutes?
- **c** Sodium-26 is a beta emitter. Write the nuclear equation for its decay.

29 A nuclear scientist has prepared equal quantities of two radioisotopes of bismuth, ^{211}Bi and ^{215}Bi. These isotopes have half-lives of 2 minutes and 8 minutes respectively. Assume when answering these questions that each sample has the same number of atoms.

- **a** Which one of the following statements best describes the activities of these samples?
  - **A** The samples start with an equal activity, then bismuth-211 has the greater activity.
  - **B** Bismuth-211 initially has four times the activity of bismuth-215.
  - **C** Bismuth-215 initially has four times the activity of bismuth-211.
  - **D** Bismuth-211 initially has twice the activity of bismuth-215.
- **b** How will the activity of these samples compare after 8 minutes?

The following information relates to questions 30 to 33. The following diagram indicates the decay series for ^{238}U through the various decay products until the stable isotope ^{206}Pb is reached.



30 Complete the nuclear reactions:

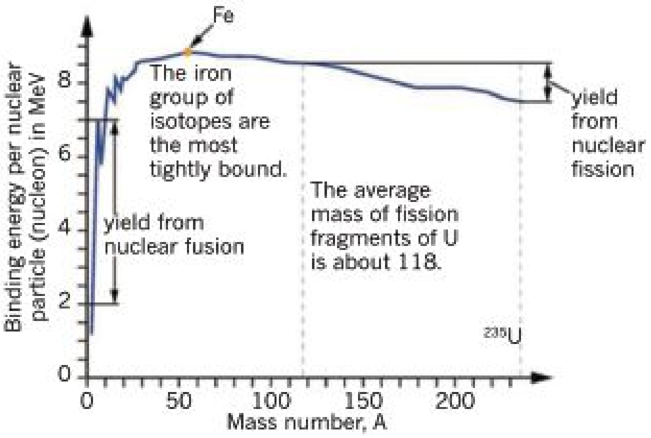
- **a** $^{238}U \rightarrow ^{234}Th +$ ______
- **b** $^{218}Po \rightarrow ^{218}At +$ ______

31 Explain why both Po and At can have the same mass number, but still be distinct elements.

32 Describe the two different decay pathways from $^{210}Bi \rightarrow ^{206}Pb$.

33 Explain the difference between the three different polonium isotopes on the chart.

The following curve should be used to answer questions 34 to 39.



34 Explain what is meant by binding energy per nucleon.

35 The element Fe is considered to be the most stable nucleus. Explain why this is the case.

36 Fission of $^{235}_{92}U$ results in fission fragments of average mass number around 118. By referring to the binding energy per nucleon for the fuel and the fragments, explain why there is a net energy release in a fission reaction.

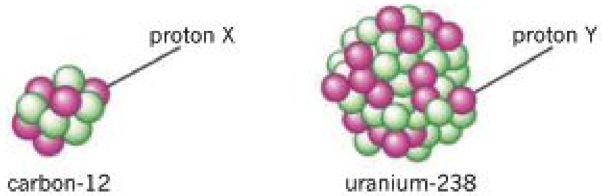
37 Estimate the energy yield of the fission of a ^{235}U nucleus to form two fission fragments of mass approximately 118.

38 Hydrogen is taken to have a binding energy per nucleon of 0 MeV. Why is this the case?

39 $^{4}_{2}He$ has a binding energy per nucleon of 7 MeV. Explain why fusion of hydrogen to form helium releases energy.

40 One kilogram of uranium-235 is capable of releasing 6.8×10^{13} J of energy during nuclear fission. In comparison, the burning of 1 kg of coal releases about 2.5×10^{7} J. Calculate the number of tonnes of coal that is burnt to provide energy equivalent to that released by 1 kg of uranium-235.

41 The diagrams below show the nuclei of carbon-12 and uranium-238. One of the protons in each nucleus is labelled.



- **a** Discuss the proportion of protons and neutrons in a small stable nuclide such as carbon-12.
- **b** Discuss the proportion of protons and neutrons in a large unstable nuclide such as uranium-238.
- **c** Discuss the sizes of the electrostatic and strong nuclear forces that are acting on proton X.
- **d** Discuss the sizes of the electrostatic and strong nuclear forces that are acting on proton Y. How do these compare with the forces acting on proton X?

42 During one of the fusion reactions that is taking place in the Sun, two helium-3 nuclei fuse together to form helium-4 and two protons.

- **a** How does the mass of the reactants compare qualitatively with that of the fusion products?
- **b** Energy is released in this reaction. What form does it take?
- **c** Discuss the forces that are acting as the two helium-3 nuclei approach each other and fuse together.

43 Consider this fission reaction of uranium-235.

$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{141}_{55}Cs + ^{93}_{37}Rb + 2^{1}_{0}n$

During this fission reaction there is a mass defect of 4.99×10^{-28} kg.

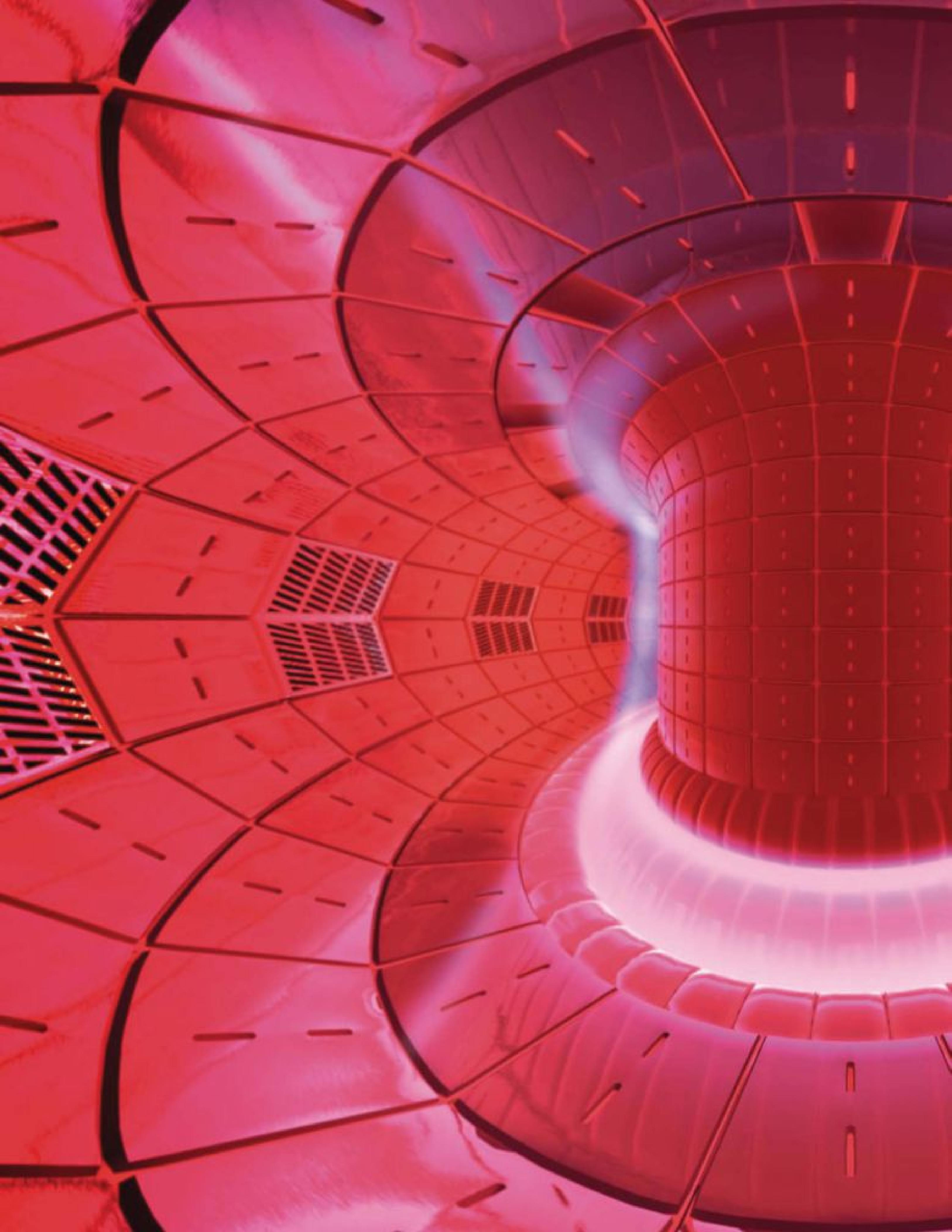
- **a** How much energy (in joules) is released by this fission?
- **b** What is this energy release in MeV?

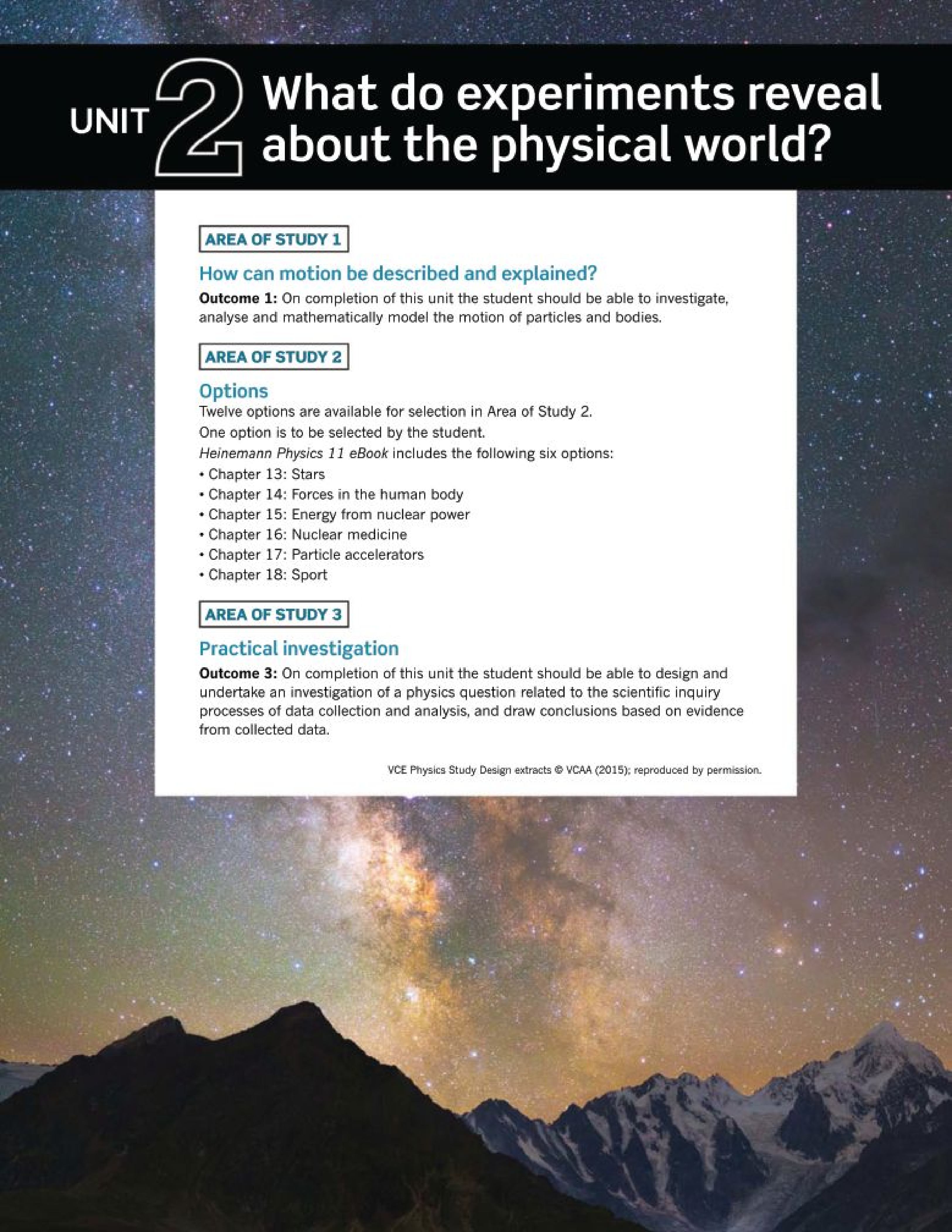
44
- **a** Two slow-moving protons are travelling directly towards each other. Will the protons collide and fuse together? Answer this question by making reference to the forces acting on the protons and the energy barrier.
- **b** Two fast-moving protons are travelling directly towards each other. The protons collide and fuse together. Discuss the forces that act on the protons and make reference to the energy barrier in your answer.

45 Fusion and fission are both processes by which energy is released from the nucleus of an atom. Identify some of the similarities and differences between these two nuclear processes.

46 A fundamental principle is that accelerating charges release radiation. Explain how and where charges are accelerated to produce the following radiation:

- **a** visible light from a tungsten filament lamp
- **b** X-rays for a medical procedure
- **c** sodium vapour lamp
- **d** red-hot piece of steel
- **e** synchrotron radiation
- **f** laser light.





UNIT 2 What do experiments reveal about the physical world?

AREA OF STUDY 1

How can motion be described and explained?

Outcome 1: On completion of this unit the student should be able to investigate, analyse and mathematically model the motion of particles and bodies.

AREA OF STUDY 2

Options

Twelve options are available for selection in Area of Study 2.

One option is to be selected by the student.

Heinemann Physics 11 eBook includes the following six options:

- Chapter 13: Stars
- Chapter 14: Forces in the human body
- Chapter 15: Energy from nuclear power
- Chapter 16: Nuclear medicine
- Chapter 17: Particle accelerators
- Chapter 18: Sport

AREA OF STUDY 3

Practical investigation

Outcome 3: On completion of this unit the student should be able to design and undertake an investigation of a physics question related to the scientific inquiry processes of data collection and analysis, and draw conclusions based on evidence from collected data.

CHAPTER 08

Scalars and vectors

Scalars and vectors are mathematical representations of quantities that are used in physics. An understanding of scalars and vectors is essential to learning concepts involving forces and motion.

By the end of this chapter, you will be able to distinguish between scalar and vector quantities. You will be able to use arrows to represent vectors and then add and subtract vectors in one and two dimensions.

Key knowledge

By the end of this chapter, you will have covered material from the study of scalars and vectors and will be able to:

- identify parameters of motion as vectors or scalars
- apply the vector model of forces, including vector addition and components of forces, to readily observable forces including the force due to gravity, friction and reaction forces
- model the force due to gravity, F_g, as the force of gravity acting at the centre of mass of a body, $F_g = mg$, where g is the gravitational field strength (9.8 N kg^{-1} near the surface of Earth).

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

8.1 Scalars and vectors

You will come into contact with many physical quantities in the natural world every day. For example, time, mass and distance are all physical quantities. Each of these physical quantities has units with which to measure them. For example, seconds, kilograms and metres.

Some measurements only make sense if there is also a direction included. For example, a GPS system tells you when to turn and in which direction. Without both of these two instructions, the information is incomplete.

All physical quantities can be divided into two broad groups based on what information you need for the quantity to make sense. These groups are called **scalars** and **vectors**. Often vectors are represented by arrows. Both of these types of measures will be investigated throughout this section.

SCALARS

There are a number of properties in nature that can be measured or determined and described using only a number and unit. For example, if the time taken for a student to travel to school is measured, you need the **magnitude** (size) and the **units** in order to understand the journey. It may take 90 minutes or one and a half hours—the number is important and the units are also important.

Quantities that require magnitude and units are called scalar quantities. Scalars do not need direction.

Examples of scalar quantities are:

- time
- distance
- volume
- speed
- temperature.

VECTORS

Vector quantities require magnitude, units and direction in order to make sense.

Examples of vectors include:

- position
- displacement
- velocity
- acceleration
- force
- momentum.

These measures are discussed in more detail in the coming chapters.

VECTORS AS ARROWS

A vector is a measurement that has both a magnitude and a direction. A vector can be visually represented as an arrow.

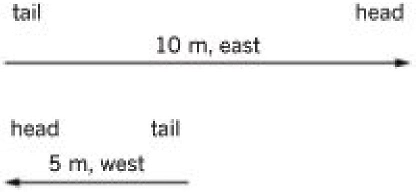


FIGURE 8.1.1 Two vector diagrams. As the top vector is twice as long as the bottom vector, it represents a measure twice the magnitude of the bottom vector. The arrowheads indicate that the vectors are in opposite directions.

Figure 8.1.1 shows two vector diagrams. In a **vector diagram,** the length of the arrow indicates the magnitude of the vector. The arrowhead shows the direction of the vector. The direction of the vector is always from its tail to its head.

A force is a push or a pull and the unit of measure for force is the newton (N). If you push a book to the right, it will respond differently to if you push the book to the left. Therefore, a force is only described properly if a direction is included, and so force is considered to be a vector. Forces are described in more detail on page 332. Force is an important concept to understand in physics, so many of the examples in this chapter refer to forces.

In most vector diagrams, the length of the arrow is drawn to scale so that it accurately represents the magnitude of the vector.

In the scaled vector diagram in Figure 8.1.2, a force $F = 4$ N left acting on the toy car is drawn as a 2 cm–long arrow. In this example a scale of 1 cm = 2 N force is used.

FIGURE 8.1.2 A force of 4 N to the left acts on a toy car.

An exact scale for the magnitude is not always used. However, it is important that vectors are drawn relative to one another. For example, a vector of 50 m north should always be about half as long as a vector of 100 m north.

Point of application of arrows

Vector diagrams may be presented slightly differently depending on what they are depicting. If the vector represents a force, the tail end of the arrow is drawn at the point where the force is applied to the object. If it is a displacement vector, attach the tail of the arrow to the position where the object starts. Friction vectors are drawn at the point where they act between an object and a surface.

Figure 8.1.3 shows a force applied by a foot to a ball (95 N east) and an opposing friction force (20 N west).

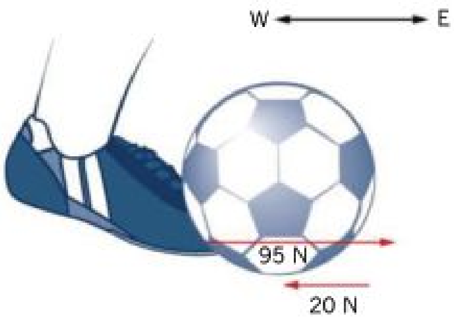


FIGURE 8.1.3 The force on the ball acts at the point of contact between the ball and the foot. The friction force acts between the ball and the ground. The kicking force, as indicated by the length of the arrow, is larger than the friction force.

DIRECTION CONVENTIONS

Vectors need a direction in order to make sense. However, for any description of a vector quantity to be useful, there needs to be a way of describing the direction that everyone understands and agrees upon.

Vectors in one dimension

For vector problems in one **dimension**, there are a number of **direction conventions** that can be used. For example:

- forwards or backwards
- up or down
- left or right.

You can also use more formal conventions such as:

- north or south
- east or west.

As you can see, for vectors in one dimension there are only two directions possible. The two directions must be in the same dimension or along the same line. The direction convention used should be presented graphically in all vector problems. This is shown in Figure 8.1.4. Arrows like these are placed near the vector diagram so that it is clear which convention is being used.

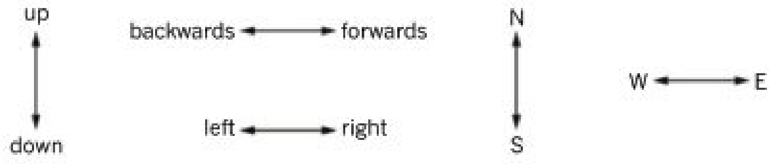


FIGURE 8.1.4 Some common one-dimensional direction conventions.

Sign convention

In calculations involving one-dimensional vectors, a sign convention can also be used to convert physical directions to the mathematical signs of positive and negative. For example, forwards can be positive and backwards can be negative, or right can be positive and left can be negative. A vector of 100 m up can be described as +100 m provided the relationship between sign and direction conventions are clearly indicated in a legend or key. Some examples are provided in Figure 8.1.5.

The advantage of using a sign convention is that the signs of positive and negative can be entered into a calculator, while the words 'up' and 'right' cannot. This is useful when adding vectors together. This will be discussed in the next section.

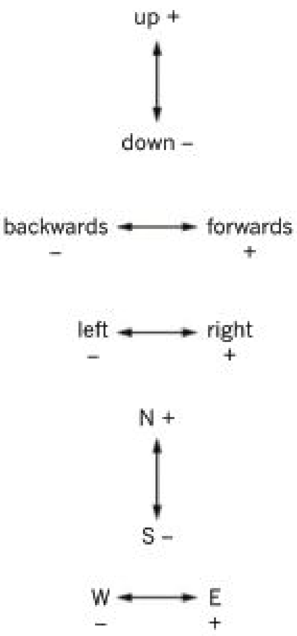
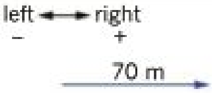


FIGURE 8.1.5 One-dimensional direction conventions can also be expressed as sign conventions.

Worked example 8.1.1

DESCRIBING VECTORS IN ONE DIMENSION



Describe the vector above using:

a the direction convention shown

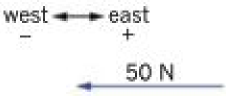
Thinking	Working
Identify the direction convention being used in the vector.	In this case the vector is pointing to the right according to the direction convention.
Note the magnitude, unit and direction of the vector.	In this example the vector is 70 m right.

b an appropriate sign convention.

Thinking	Working
Convert the physical direction to the corresponding mathematical sign.	The physical direction of right is positive and left is negative. In this example, the arrow is pointing right, so the mathematical sign is +.
Represent the vector with a mathematical sign, magnitude and unit.	This vector is +70 m.

Worked example: Try yourself 8.1.1

DESCRIBING VECTORS IN ONE DIMENSION



Describe the vector above using:

a the direction convention shown

b an appropriate sign convention.

Vectors in two dimensions

When vectors are in one dimension, it is relatively simple to understand direction. However, some vectors will require a description in a two-dimensional plane. These planes could be:

- horizontal, which can be defined using north, south, east and west
- vertical, which can be defined in a number of ways including forwards, backwards, up, down, left and right.

The description of the direction of these vectors is more complicated. Therefore, a more detailed convention is needed for identifying the direction of a vector. There are a variety of conventions, but they all describe a direction as an angle from a known reference point.

Horizontal plane

For a horizontal, two-dimensional plane, the two common methods for describing the direction of a vector are:

- full circle (or true) bearing. A 'full circle bearing' describes north as zero degrees true. This is written as 0° T. In this convention, all directions are given as a clockwise angle from north. As an example, 95° T is 95° clockwise from north.
- quadrant bearing. An alternative method is to provide a 'quadrant bearing', where all angles are referenced from either north or south and are between 0° and 90° towards east or west. In this method, 30° T becomes N 30° E, which can be read as 'from north 30° towards the east'.

Using these two conventions, north-west (NW) would be 315° T using a full circle bearing, or N 45° W using a quadrant bearing. Figure 8.1.6 demonstrates these two methods.

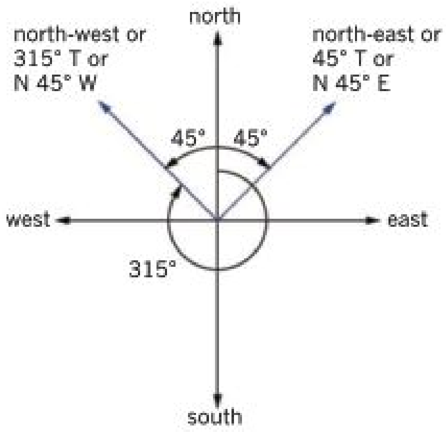
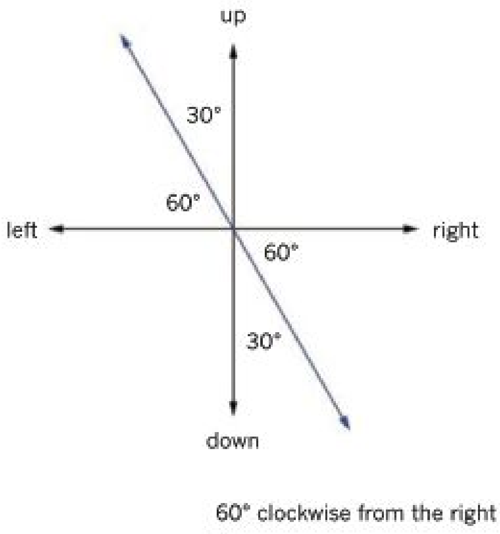


FIGURE 8.1.6 Two horizontal vector directions, viewed from above, using full circle bearings and quadrant bearings.

Vertical plane

For a vertical, two-dimensional plane the directions are referenced to vertical (upwards and downwards) or horizontal (left and right) and are between 0° and 90° clockwise or anticlockwise. For example, a vector direction can be described as '60° clockwise from the left direction'. The same vector direction could be described as '30° anticlockwise from the upwards direction'. The opposite direction to this vector would be '60° clockwise from the right direction'. This example is illustrated in Figure 8.1.7.

30° anticlockwise from the upwards direction
or
60° clockwise from the left direction



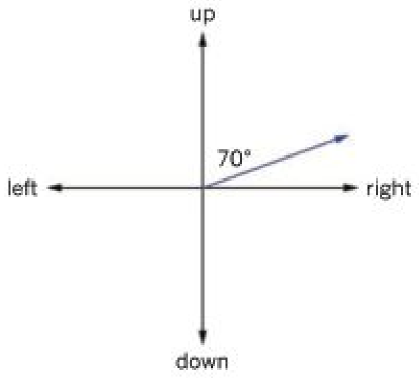
60° clockwise from the right direction
or
30° anticlockwise from the downwards direction

FIGURE 8.1.7 Two vectors in the vertical plane.

Worked example 8.1.2

DESCRIBING TWO-DIMENSIONAL VECTORS

Describe the direction of the following vector using an appropriate method.

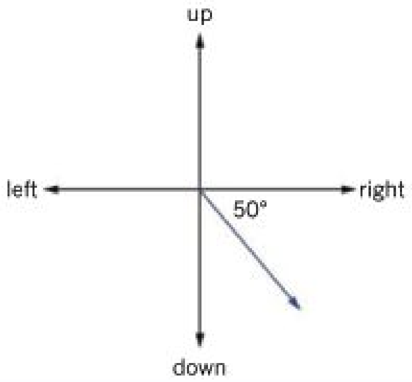


Thinking	Working
Choose the appropriate points to reference the direction of the vector. In this case using the vertical reference makes more sense, as the angle is given from the vertical.	The vector can be referenced to the vertical.
Determine the angle between the reference direction and the vector.	In this example there is 70° from vertically up to the vector.
Determine the direction of the vector from the reference direction.	From vertically up, the vector is clockwise.
Describe the vector using the sequence: angle, clockwise or anticlockwise from the reference direction.	This vector is 70° clockwise from the upwards direction.

Worked example: Try yourself 8.1.2

DESCRIBING TWO-DIMENSIONAL VECTORS

Describe the direction of the following vector using an appropriate method.



8.1 Review

SUMMARY

- Scalar quantities require a magnitude and a unit to make sense. No direction is required for scalar quantities.
- Distance, time, speed and mass are examples of scalar quantities.
- Vectors require magnitude, units and direction to make sense.
- Displacement, velocity, acceleration and force are examples of vectors.
- Arrows are used to represent vectors.
- The length of the arrow represents the magnitude of the vector.
- The direction the arrow is pointing indicates the direction of the vector.
- Vector arrows can be drawn to scale, or drawn relative to each other.
- Force vectors are drawn with their tails attached to the point of application on the object.
- Displacement vectors are drawn from the start of the journey towards the end of the journey.
- One-dimensional vectors use direction conventions and sign conventions to describe the direction of the vector. Examples include left and right, up and down, + and –.
- The direction of two-dimensional vectors in the horizontal plane can be described using a full circle bearing or a quadrant bearing. Vectors in the vertical plane can be described using angles measured clockwise and anticlockwise from the vertical or horizontal.

KEY QUESTIONS

1 What information is required to fully describe a scalar measure?

2 What information is required to fully describe a vector measure?

3 Classify each of the following as scalar or vector quantities.
  a time
  b force
  c acceleration
  d distance
  e position
  f displacement
  g volume
  h momentum
  i speed
  j velocity
  k temperature

4 For the following, decide which of the vector magnitudes provided describes which vector diagram. Note: one of the vector magnitudes is not required.
5.4 N; 2.7 N; 9.0 N; 8.1 N

(a)

(b)

(c)

5 For the following, decide which of the vector magnitudes provided describes which vector diagram. Note: one of the vector magnitudes is not required.
10.8 N; –2.7 N; –5.4 N; 16.2 N

(a)

(b)

(c)

6 Give the opposing direction to each of the following one-dimensional descriptions.
  a up
  b north
  c backwards
  d down
  e west
  f negative

7 Why is it sometimes appropriate to rename direction conventions with a positive or negative sign—for example, + instead of north or – instead of left?

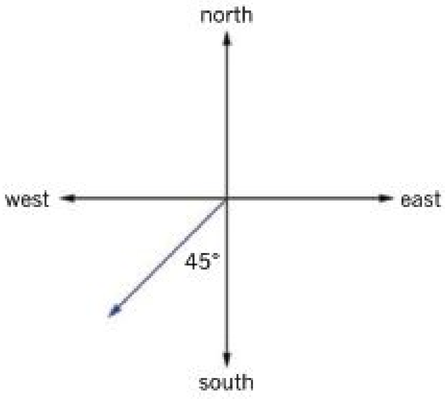
8 Describe the following vector using an appropriate convention.

8.1 Review *continued*

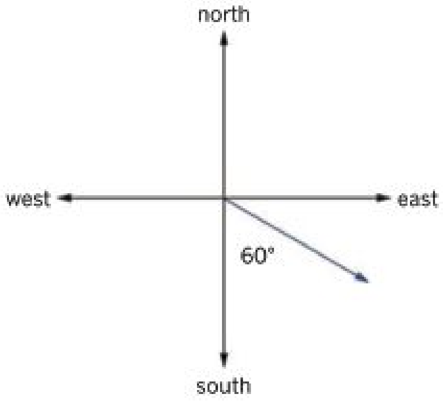
9 Describe the following vectors using:

- i full circle bearings
- ii quadrant bearings.

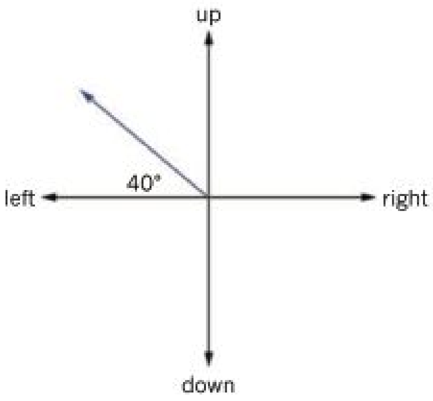
a



b



10 Describe the following vector using appropriate conventions.



8.2 Adding vectors in one and two dimensions

In real situations, more than one vector may act on an object. If this is the case, it is helpful to analyse the associated vector diagrams to find out the overall or combined effect of the vectors.

When vectors are combined, it is called adding vectors. Adding vectors that are in one dimension means finding the resultant vector for a number of vectors that are in the same line. It is also possible to find the resultant of a number of vectors that are in two dimensions. The vectors can be in any direction, as long as they are all in the same plane.

ADDING VECTORS IN ONE DIMENSION

When two or more vectors are in the same dimension, it means that the vectors are either pointing in the same direction or in the opposite direction. They are **collinear** (in line with each other). For example, the vector measurements 10 m west, 15 m east and 25 m west are all in one dimension. They are all in the same or opposite direction to each other.

Graphical method of adding vectors

Vector diagrams, like those shown in Figure 8.2.1, are convenient for adding vectors. To combine vectors in one dimension, draw the first vector, then start the second vector with its tail at the head of the first vector. Continue adding arrows 'head to tail' until the last vector is drawn. The sum of the vectors, or the **resultant** vector, is drawn from the tail of the first vector to the head of the last vector.

Vectors are added head to tail. The resultant vector is drawn from the tail of the first vector to the head of the last vector.

FIGURE 8.2.1 Adding vectors head to tail. This particular diagram represents the addition of 15 m east and 5 m east. The resultant vector, shown in red, is 20 m east.

In Figure 8.2.1, the two vectors s_1 (15 m east) and s_2 (5 m east) are drawn separately. The two vectors are then drawn with s_1 and s_2 connected head to tail. The resultant vector s_R is drawn from the tail of s_1 to the head of s_2. The magnitude (size) of the resultant vector can be deduced from the magnitudes of the separate vectors: 15 m + 5 m = 20 m.

Alternatively, vectors can be drawn to scale, for example: 1 m = 1 cm. The resultant vector is then directly measured from the scale diagram. The direction of the resultant vector is the same as the direction from the tail of the first vector to the head of the last vector.

Algebraic method of adding vectors

To add vectors in one dimension algebraically, a sign convention is used to represent the direction of the vectors (see Figure 8.2.2). When applying a sign convention, it is important to provide a key explaining the convention used.

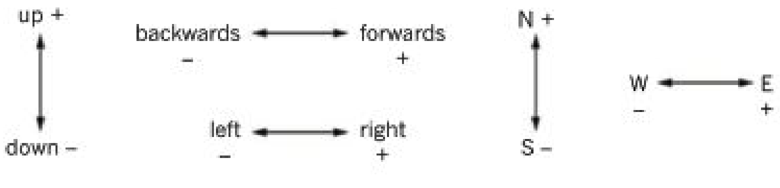


FIGURE 8.2.2 Common sign and direction conventions.

The sign convention allows you to enter the signs and magnitudes of vectors into a calculator. The sign of the final magnitude gives the direction of the resultant vector.

Worked example 8.2.1

ADDING VECTORS IN ONE DIMENSION USING ALGEBRA

Use the sign and direction conventions shown in Figure 8.2.2 to determine the resultant vector of a student who walks 25 m west, 16 m east, 44 m west and then 12 m east.	
Thinking	**Working**
Apply the sign conventions to change each of the directions to signs.	25 m west = –25 m 16 m east = +16 m 44 m west = –44 m 12 m east = +12 m
Add the magnitudes and their signs together.	Resultant vector = (–25) + (+16) + (–44) + (+12) = –41 m
Refer to the sign and direction conventions to determine the direction of the resultant vector.	Negative is west. ∴ Resultant vector = 41 m west

Worked example: Try yourself 8.2.1

ADDING VECTORS IN ONE DIMENSION USING ALGEBRA

Use the sign and direction conventions shown in Figure 8.2.2 on page 265 to determine the resultant force on a box that has the following forces acting on it: 16 N up, 22 N down, 4 N up and 17 N down.

ADDING VECTORS IN TWO DIMENSIONS

Adding vectors in two dimensions means that all of the vectors must be in the same plane. The vectors can go in any direction within the plane, and can be separated by any angle. The examples in this section illustrate vectors in the horizontal plane, but the same strategies apply to adding vectors in the vertical plane.

The horizontal plane is one that is looked down on from above. Examples include looking at a house plan or map placed on a desk. The direction conventions that suit this plane best are the north, south, east and west convention, or the forwards, backwards, left and right convention. These are shown in Figure 8.2.3.

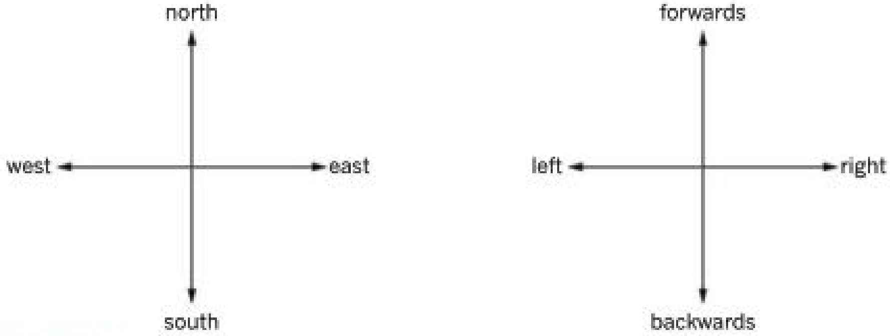


FIGURE 8.2.3 The direction conventions for the horizontal plane.

When two vectors are in the horizontal plane, the angles between them can be right-angled, acute or obtuse.

Graphical method of adding vectors

The magnitude and direction of a resultant vector can be determined by measuring an accurately drawn scaled vector diagram. There are two main ways to do this:

- head to tail method
- parallelogram method.

Head to tail method

To add vectors at right angles to each other using a graphical method, use an appropriate scale and then draw each vector head to tail. The resultant vector is the vector that starts at the tail of the first vector and ends at the head of the last vector. To determine the magnitude and direction of the resultant vector, measure the length of the resultant vector and compare it to the scale, then measure and describe the direction appropriately.

In Figure 8.2.4, two vectors, 30.0 m east and 20.0 m south, are added head to tail. The resultant vector, shown in red, is measured to be about 36 m according to the scale provided. Using a protractor, the resultant vector is measured to be in the direction 34° south of east. This represents a direction of S 56° E when using quadrant bearings.

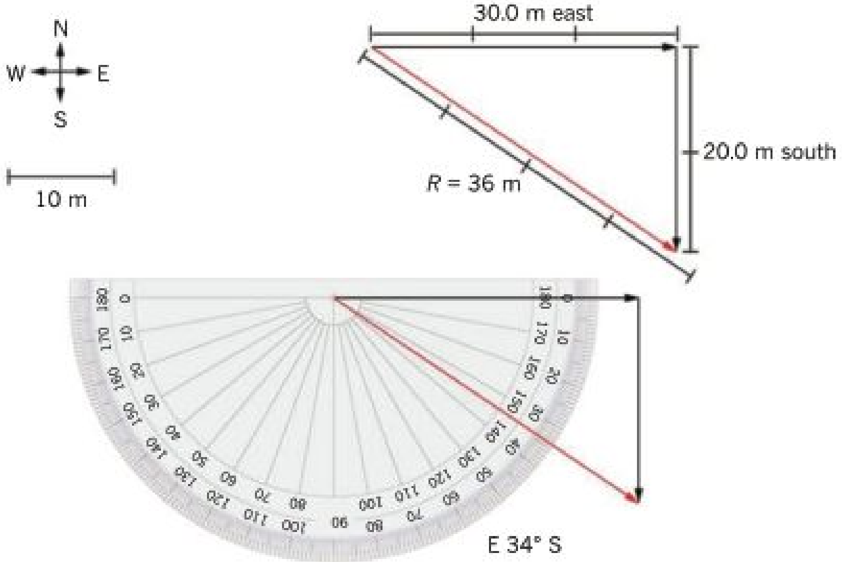


FIGURE 8.2.4 Adding two vectors at right angles, using the graphical method.

If the two vectors are at angles other than 90° to each other, the graphical method is ideal for finding the resultant vector. In Figure 8.2.5, the vectors 15 N east and 10 N S 45° E are added head to tail. The magnitude of the resultant vector is measured to be about 23 N. The direction of the resultant vector is measured by a protractor from east to be 18° towards the south, which should be written as S 72° E.

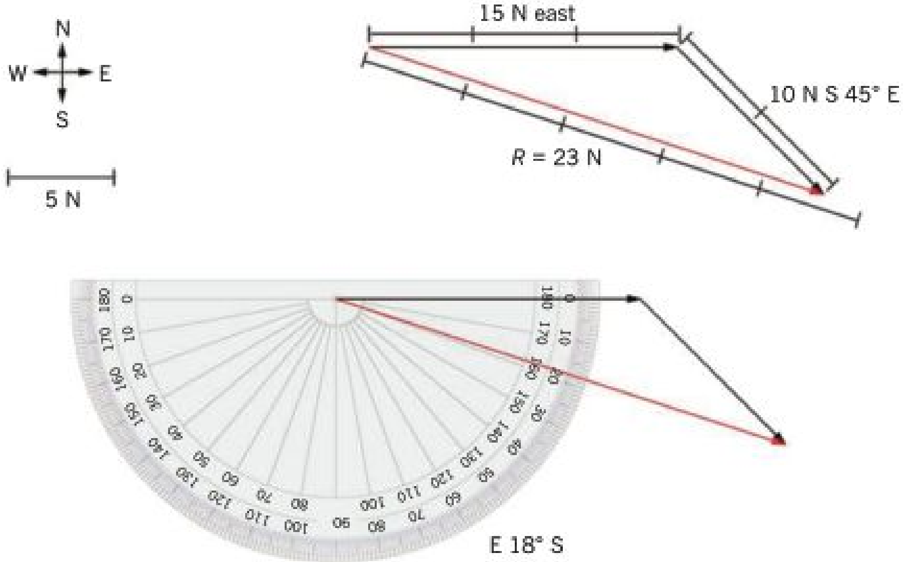


FIGURE 8.2.5 Adding two vectors not at right angles, using the graphical method.

Parallelogram method

An alternative method for determining a resultant vector is to construct a parallelogram of vectors. In this method the two vectors to be added are drawn tail to tail. Next, a parallel line is drawn for each vector as shown in Figure 8.2.6. In this figure, the parallel lines have been drawn as dotted lines. The resultant vector is drawn from the tails of the two vectors to the intersection of the dotted parallel lines.

FIGURE 8.2.6 Parallelogram of vectors method for adding two vectors.

PHYSICSFILE

Trigonometric ratios

Most students learn the mnemonic SOHCAHTOA in their maths classes. It is often pronounced soh-cah-toa and provides a way to remember the trigonometric ratios:

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$$

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$

Geometric method of adding vectors

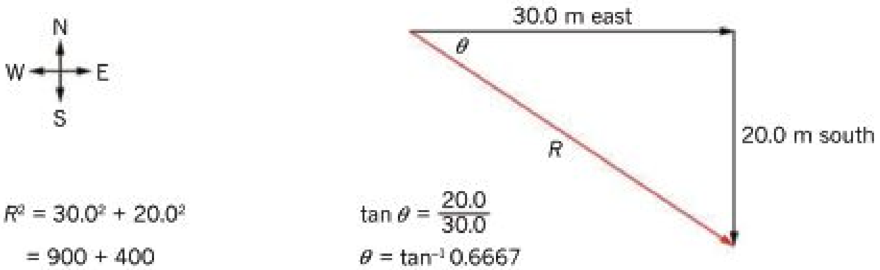
Graphical methods of adding vectors in two dimensions only give approximate results as they rely on comparing the magnitude of the resultant vector to a scale and measuring the direction with a protractor. A more accurate method to resolve vectors is to use Pythagoras' theorem and trigonometry. These techniques are referred to as geometric methods. Geometric methods can be used to calculate the magnitude of the vector and its direction. However, Pythagoras' theorem and trigonometry can only be used for finding the resultant vector of two vectors that are at right angles to each other.

PHYSICSFILE

Pythagoras' theorem

Pythagoras' theorem is $a^2 + b^2 = c^2$ where c is the hypotenuse (the longest side) and a and b are the two shorter sides of a right-angled triangle. The hypotenuse is easily recognised as it is directly across from (opposite) the right angle of the triangle.

In Figure 8.2.7, two vectors, 30.0 m east and 20.0 m south, are added head to tail. The resultant vector, shown in red, is calculated using Pythagoras' theorem to be 36.1 m. The resultant vector is calculated to be in the direction S 56.3° E. This result is more accurate than the answer determined earlier in this section.



$$\begin{aligned} R^2 &= 30.0^2 + 20.0^2 \\ &= 900 + 400 \\ R &= \sqrt{1300} \\ &= 36.1\text{ m} \end{aligned}$$

$$\begin{aligned} \tan\theta &= \frac{20.0}{30.0} \\ \theta &= \tan^{-1} 0.6667 \\ &= 33.7° \end{aligned}$$

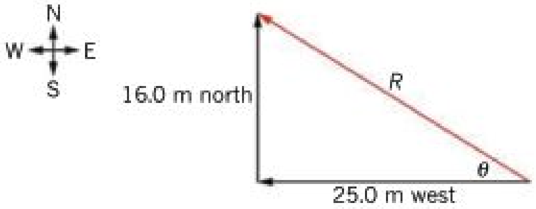
E 33.7° S (or S 56.3° E)

FIGURE 8.2.7 Adding two vectors at right angles, using the geometric method.

Worked example 8.2.2

ADDING VECTORS IN TWO DIMENSIONS USING GEOMETRY

Determine the resultant vector that represents a child running 25.0 m west and 16.0 m north. Refer to Figure 8.2.2 on page 265 for sign and direction conventions if required.

Thinking	**Working**
Construct a vector diagram showing the vectors drawn head to tail. Draw the resultant vector from the tail of the first vector to the head of the last vector.	
As the two vectors to be added are at 90° to each other, apply Pythagoras' theorem to calculate the magnitude of the resultant vector.	$R^2 = 25.0^2 + 16.0^2$ $= 625 + 256$ $R = \sqrt{881}$ $= 29.7$ m
Using trigonometry, calculate the angle from the west vector to the resultant vector.	$\tan\theta = \frac{16.0}{25.0}$ $\theta = \tan^{-1} 0.640$ $= 32.6°$
Determine the direction of the vector relative to north or south.	$90° - 32.6° = 57.4°$ The direction is N 57.4° W
State the magnitude and direction of the resultant vector.	$R = 29.7$ m, N 57.4° W

Worked example: Try yourself 8.2.2

ADDING VECTORS IN TWO DIMENSIONS USING GEOMETRY

Determine the resultant force when forces of 5.0 N east and 3.0 N north act on a tree. Refer to Figure 8.2.2 for sign and direction conventions if required.

PHYSICS IN ACTION

Surveying

Surveyors use technology to measure, analyse and manage data about the shape of the land and the exact location of landmarks and buildings. They take many measurements, including angles and distances, and use them to calculate more advanced data such as vectors, bearings, co-ordinates, elevations, maps etc. Surveyors typically use theodolites (see Figure 8.2.8 and Figure 8.2.9), GPS survey equipment, laser range finders and satellite images to map the land in three dimensions.

FIGURE 8.2.8 Surveying the land with a theodolite.

Surveyors are often the first professionals on a building site to ensure that the boundaries of the property are correct. They also ensure that the building is built in the correct location. Surveyors must liaise closely with architects both before and during a building project as they provide position and height data for walls and floors.

FIGURE 8.2.9 Surveying equipment being used on a building site.

8.2 Review

SUMMARY

- Combining vectors is known as adding vectors.
- One-dimensional vector addition refers to vectors in a line, while two-dimensional vector addition refers to vectors on a plane.
- Adding vectors in one dimension can be done graphically by using vector diagrams. After adding vectors head to tail, the resultant vector can be drawn from the tail of the first vector to the head of the last vector.
- Adding vectors in one dimension can be done algebraically by applying a sign convention. Vectors with direction become vectors with either positive or negative signs.
- Adding vectors in two dimensions can be estimated graphically with a scale and a protractor.
- An alternate method of adding vectors in two dimensions is to construct a parallelogram of vectors.
- Adding vectors in two dimensions can be calculated using Pythagoras' theorem and the trigonometric ratios of a right-angled triangle.

KEY QUESTIONS

1 Find the resultant vector when the following are combined: 2 m west, 5 m east and 7 m west.

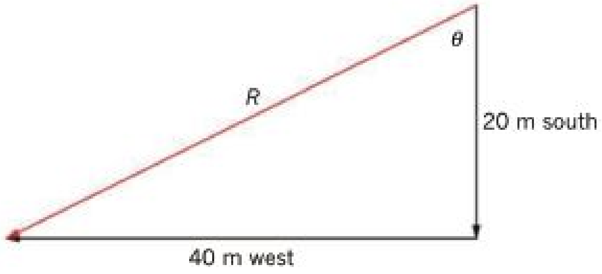
2 Add the following vectors to find the resultant vector: 3 m up, 2 m down and 3 m down.

3 Determine the resultant vector of a toy train that is made to move in these directions: 23 m forwards, 16 m backwards, 7 m forwards and 3 m backwards.

4 When adding vector B to vector A using the head to tail method, from what point, and to what point, is the resultant vector drawn?

A from the head of A, to the tail of B
B from the tail of B, to the head of A
C from the head of B, to the tail of A
D from the tail of A, to the head of B

5 Describe the magnitude and direction of the resultant vector, drawn in red, in the following diagram.



6 Forces of 2000 N north and 6000 N east act on an object. What is the resultant force acting?

7 What is the magnitude of the resultant vector when 30.0 m south and 40.0 m west are added?

A 7.7 m
B 44.7 m
C 50.0 m
D 2000 m

8.3 Subtracting vectors in one and two dimensions

The previous section discussed combining or adding vectors. In physics there are times when the difference between two vectors has to be determined. For example, a change in velocity is determined by the final velocity minus the initial velocity. In other words, you must subtract vectors. One way to subtract one vector from another is to add the opposite vector.

SUBTRACTING VECTORS IN ONE DIMENSION

To find the difference between two vectors, you must subtract the initial vector from the final vector. To do this, work out which is the initial vector, then reverse its direction. These two vectors are then added: the final vector and the opposite of the initial vector.

This technique can be applied both graphically and algebraically.

PHYSICSFILE

Double negatives

When a negative number is multiplied by another negative number, the result is a positive number. It is also illustrated when a negative number is subtracted from another number. The effect is to add the two numbers together. For example, $(5) - (-2) = 7$.

It is important to differentiate between the terms subtract, minus, take away or difference between and the term negative. The terms subtract, minus, take away or difference between are processes, like add, multiply and divide. You will find them grouped together on your calculator. The term negative is a property of a number that means that it is opposite to positive. There is a separate button on your calculator for this property.

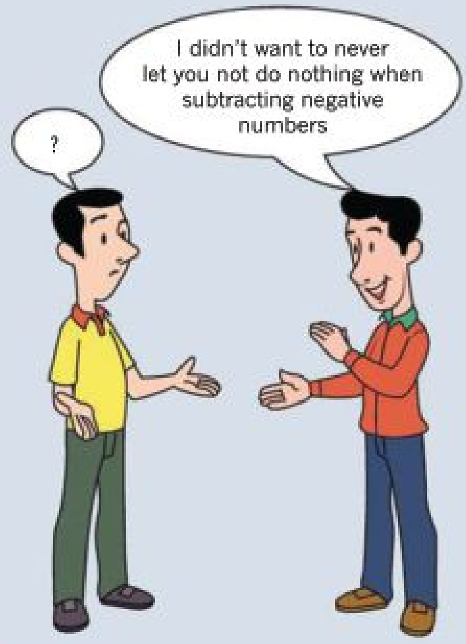


FIGURE 8.3.1

Graphical method of subtracting vectors

Velocity is a quantity that gives an indication of how fast an object is moving. It is a vector because the direction is important when stating the velocity of an object. For example, the velocity of the tennis ball moving towards the racquet in Figure 8.3.2 on page 273 is different from the velocity of the tennis ball as it leaves the racquet. The concept of velocity is covered in more detail in Chapter 9, but it is useful to use the example of velocity now when discussing the subtraction of vectors. The processes applied to the subtraction of velocity vectors works for all other vectors.

To subtract velocity vectors in one dimension using a graphical method, determine which vector is the initial velocity and which is the final velocity. The final velocity is drawn first. The initial velocity is then drawn, but in the opposite direction to its original form. The sum of these vectors, or the resultant vector, is drawn from the tail of the final velocity to the head of the reversed initial velocity. This resultant vector is the difference between the two velocities, or Δv.

FIGURE 8.3.2 As velocity is a vector, direction is important. The tennis ball has a different velocity when it leaves the racquet from when it travelled towards the racquet.

To find the difference betweena or change in vectors, subtract the initial vector from the final vector. Vectors are subtracted by adding the negative of a vector.

In Figure 8.3.3, the two separate velocity vectors, v_1 (9 m s^{-1} east) and v_2 (3 m s^{-1}, east) are drawn separately. The initial velocity, v_1, is then drawn again in the opposite direction: $-v_1$ or 9 m s^{-1} west.

FIGURE 8.3.3 Subtracting vectors using the graphical method.

Figure 8.3.4 illustrates how the difference between the vectors is found. Firstly, the final velocity, v_2, is drawn. Then the opposite of the initial velocity, $-v_1$, is drawn head to tail. The resultant vector, Δv, is drawn from the tail of v_2 to the head of $-v_1$.

FIGURE 8.3.4 Subtracting vectors using the graphical method.

The magnitude of the resultant vector, Δv, can be calculated from the magnitudes of the two vectors. Alternatively, you could draw the vectors to scale and then measure the resultant vector against that scale, for example 1 m s^{-1} = 1 cm.

The direction of the resultant vector, Δv, is the same as the direction from the tail of the final velocity, v_2, to the head of the opposite of the initial velocity, $-v_1$.

Algebraic method of subtracting vectors

To subtract velocity vectors in one dimension algebraically, a sign convention is used to represent the direction of the velocities. Some examples of one-dimensional directions include east and west, north and south and up and down. These options are replaced by positive (+) or negative (–) signs when calculations are performed. To change the direction of the initial velocity, simply change the sign from positive to negative or from negative to positive.

The equation for finding the change in velocity is:

change in velocity = final velocity – initial velocity

$\Delta v = v_2 - v_1$

$\Delta v = v_2 + (-v_1)$

change in velocity = final velocity + the opposite of the initial velocity

The final velocity is added to the opposite of the initial velocity. Since the change in velocity is a vector, it will consist of a sign and a magnitude. The sign of the answer can be compared with the sign and direction convention (Figure 8.3.5) to determine the direction of the change in velocity.

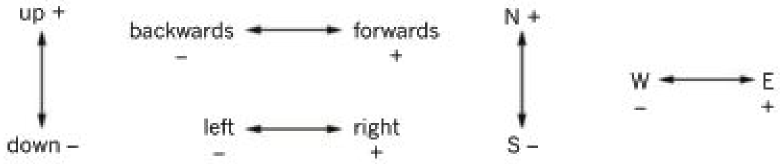


FIGURE 8.3.5 Sign and direction conventions.

Worked example 8.3.1

SUBTRACTING VECTORS IN ONE DIMENSION USING ALGEBRA

Use the sign and direction conventions shown in Figure 8.3.5 to determine the change in velocity of a plane as it changes from 255 m s^{-1} west to 160 m s^{-1} east.

Thinking	Working
Apply the sign and direction convention to change the directions to signs.	v_1 = 255 m s^{-1} west = −255 m s^{-1} v_2 = 160 m s^{-1} east = +160 m s^{-1}
Reverse the direction of the initial velocity, v_1, by reversing the sign.	$-v_1$ = 255 m s^{-1} east = +255 m s^{-1}
Use the formula for change in velocity to calculate the magnitude and the sign of Δv.	$\Delta v = v_2 + (-v_1)$ = (+160) + (+255) = +415 m s^{-1}
Refer to the sign and direction convention to determine the direction of the change in velocity.	Positive is east ∴ Δv = 415 m s^{-1} east

Worked example: Try yourself 8.3.1

SUBTRACTING VECTORS IN ONE DIMENSION USING ALGEBRA

Use the sign and direction conventions shown in Figure 8.3.5 to determine the change in velocity of a rocket as it changes from 212 m s^{-1} up to 2200 m s^{-1} up.

SUBTRACTING VECTORS IN TWO DIMENSIONS

Changing velocity in two dimensions can occur when turning a corner. For example, walking at 3 m s^{-1} west, then turning to travel at 3 m s^{-1} north. Although the magnitude of the velocity is the same, the direction is different.

When a change in a vector occurs, the magnitude and/or the direction of the vector can change.

A change in velocity in two dimensions can be determined using either the graphical method or the geometric method described in the previous section. The initial velocity must always be reversed before it is added to the final velocity.

The two dimensional direction conventions were introduced in the previous section and are shown here in Figure 8.3.6.

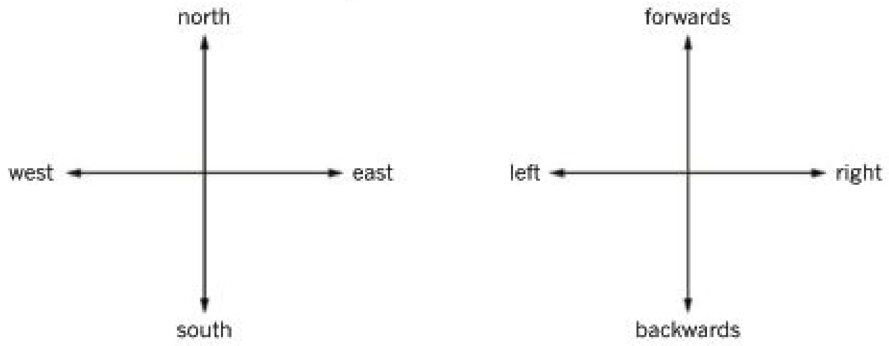


FIGURE 8.3.6 The direction conventions for the horizontal plane.

Graphical method of subtracting vectors

To subtract vectors using a graphical method, use a direction convention and a scale and draw each vector.

Using velocity as an example, the steps to do this are as follows:

- Draw in the final velocity first.
- Draw the opposite of the initial velocity head to tail with the final velocity vector.
- Draw the resultant change in velocity vector, starting at the tail of the final velocity vector and ending at the head of the opposite of the initial velocity vector.
- Measure the length of the resultant vector and compare it to the scale to determine the magnitude of the change in velocity.
- Measure an appropriate angle to determine the direction of the resultant vector.

Figure 8.3.7 shows the velocity vectors for travelling 3 m s^{-1} west and then turning and travelling 3 m s^{-1} north. The opposite of the initial velocity is drawn as 3 m s^{-1} east.

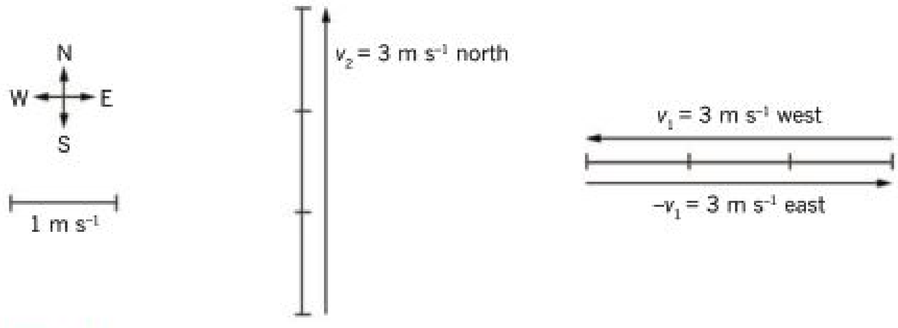


FIGURE 8.3.7 Subtracting two vectors at right angles, using the graphical method.

To determine the change in velocity, the final velocity vector is drawn first, then from its head the opposite of the initial velocity is drawn. This is shown in Figure 8.3.8. The magnitude of the change in velocity (resultant vector) is shown in red. It is measured to be about 4.3 m s^{-1} according to the scale provided. Using a protractor, the resultant vector is measured to be in the direction N 45° E.

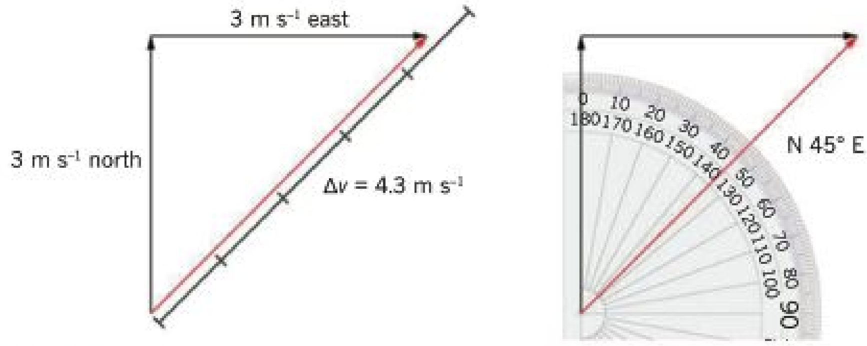


FIGURE 8.3.8 Subtracting two vectors at right angles, using the graphical method.

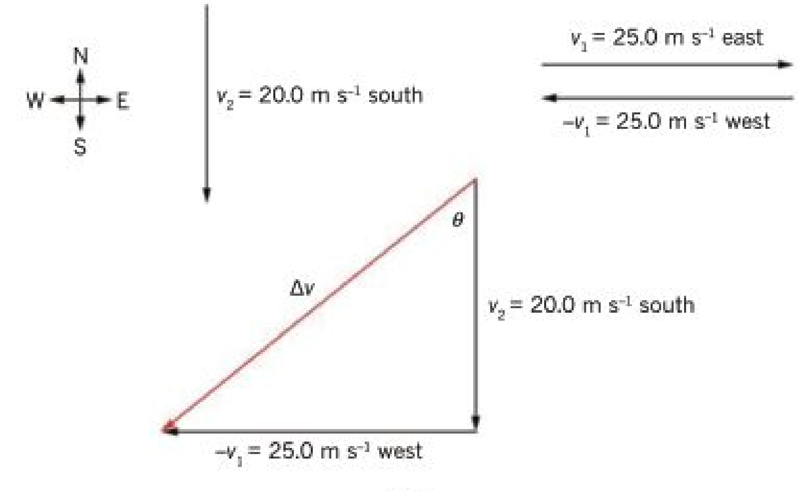
Geometric method of subtracting vectors

The graphical method of subtracting vectors in two dimensions only gives approximate results, as it relies on comparing the magnitude of the change in velocity vector to a scale and measuring its direction with a protractor.

A more accurate method to subtract vectors is to use Pythagoras' theorem and trigonometry.

Figure 8.3.9 shows how to calculate the resultant velocity when changing from 25 m s^{-1} east to 20.0 m s^{-1} south. The initial velocity of 25.0 m s^{-1} east and the final velocity of 20.0 m s^{-1} south are drawn. Then the opposite of the initial velocity is drawn as 25.0 m s^{-1} west. The final velocity vector is drawn first, then from its head the opposite of the initial velocity is drawn. The resultant velocity vector, shown in red, is calculated to be 32.0 m s^{-1}. The resultant vector is calculated to be in the direction S 51.3° W.

The resultant vector is 32.0 m s^{-1} S 51.3° W.



$R^2 = 25.0^2 + 20.0^2$

$= 625 + 400$

$R = \sqrt{1025}$

$= 32.0$ m s^{-1}

$\tan\theta = \dfrac{25.0}{20.0}$

$\theta = \tan^{-1} 1.25$

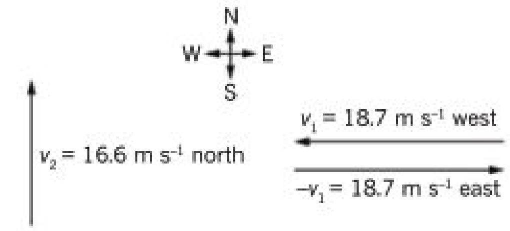
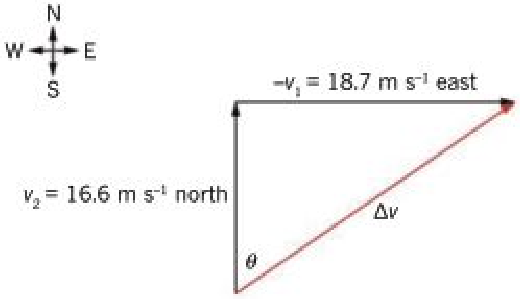
$= 51.3°$

S 51.3° W

FIGURE 8.3.9 Subtracting two vectors at right angles, using the geometric method.

Worked example 8.3.2

SUBTRACTING VECTORS IN TWO DIMENSIONS USING GEOMETRY

<table>
<tr><td colspan="2">Determine the change in velocity of Clare's scooter as she turns a corner if she approaches it at 18.7 m s^{-1} west and exits at 16.6 m s^{-1} north.</td></tr>
<tr><td>Thinking</td><td>Working</td></tr>
<tr><td>Draw the final velocity vector, v_2, and the initial velocity vector, v_1, separately. Then draw the initial velocity in the opposite direction.</td><td>
</td></tr>
<tr><td>Construct a vector diagram drawing v_2 first and then from its head draw the opposite of v_1. The change of velocity vector is drawn from the tail of the final velocity to the head of the opposite of the initial velocity.</td><td>
</td></tr>
<tr><td>As the two vectors to be added are at 90° to each other, apply Pythagoras' theorem to calculate the magnitude of the change in velocity.</td><td>$R^2 = 16.6^2 + 18.7^2$
$= 275.26 + 349.69$
$R = \sqrt{625.25}$
$= 25.0 \text{ m s}^{-1}$</td></tr>
<tr><td>Calculate the angle from the north vector to the change in velocity vector.</td><td>$\tan \theta = \frac{18.7}{16.6}$
$\theta = \tan^{-1} 1.16$
$= 48.4°$</td></tr>
<tr><td>State the magnitude and direction of the change in velocity.</td><td>$\Delta v = 25.0 \text{ m s}^{-1}$ N 48.4° E</td></tr>
</table>

Worked example: Try yourself 8.3.2

SUBTRACTING VECTORS IN TWO DIMENSIONS USING GEOMETRY

Determine the change in velocity of a ball as it bounces off a wall. The ball approaches at 7.0 m s^{-1} south and rebounds at 6.0 m s^{-1} east.

8.3 Review

SUMMARY

- To find the difference between, or change in vectors, subtract the initial vector from the final vector.
- Vectors are subtracted by adding the negative, or opposite, of a vector.
- Vector subtraction in one or two dimensions can be determined graphically using a scale and a protractor.
- Vector subtraction in one dimension can be determined algebraically.
- Vector subtraction in two dimensions can be determined geometrically using Pythagoras' theorem and trigonometry.

KEY QUESTIONS

1. A car that was initially travelling at a velocity of 3 m s^{-1} west is later travelling at 5 m s^{-1} east. What is the difference between the two vectors?
2. Determine the change in velocity of a runner who changes from running at 4 m s^{-1} to the right on grass to running 2 m s^{-1} to the right in sand.
3. A student throws a ball up into the air at 4 m s^{-1}. A short time later the ball is travelling back downwards to hit the ground at 3 m s^{-1}. Determine the change in velocity of the ball during this time.
4. Tom hits a tennis ball against a wall. If the ball travels towards the wall at 35.0 m s^{-1} north and rebounds at 32.5 m s^{-1} south, calculate the change in velocity of the ball.
5. Jamelia applies the brakes on her car and changes her velocity from 22.2 m s^{-1} forwards to 8.20 m s^{-1} forwards. Calculate the change in velocity of Jamelia's car.
6. A jet plane makes a turn after taking off, changing its velocity from 345 m s^{-1} south to 406 m s^{-1} west. Calculate the change in the velocity of the jet.
7. Yvette hits a golf ball that strikes a tree and changes its velocity from 42.0 m s^{-1} east to 42.0 m s^{-1} north. Calculate the change in the velocity of the golf ball.
8. A yacht tacks during a race, changing its velocity from 7.05 m s^{-1} south to 5.25 m s^{-1} west. Calculate the change in the velocity of the yacht.

8.4 Vector components

Sections 8.2 and 8.3 explored how vectors can be combined to find a resultant vector. In physics there are times when it is useful to break one vector up into two vectors that are at right angles to each other. For example, if a force vector is acting at an angle up from horizontal, as shown in Figure 8.4.1, this vector can be considered to consist of two independent vertical and horizontal components.

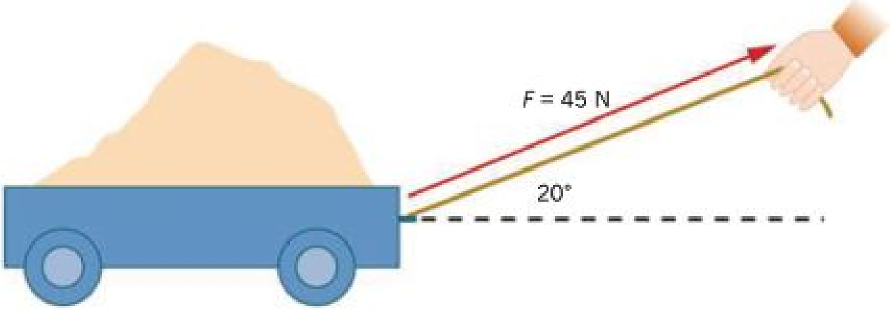


FIGURE 8.4.1 The pulling force acting on the cart has a component in the horizontal direction and a component in the vertical direction.

The components of a vector can be found using trigonometry.

FINDING PERPENDICULAR COMPONENTS OF A VECTOR

Vectors at an angle are more easily dealt with if they are broken up into perpendicular **components**, that is, two components that are at right angles to each other. These components, when added together, give the original vector. To find the components of a vector, a right-angled triangle is constructed with the original vector as the hypotenuse. This is shown in Figure 8.4.2. The hypotenuse is always the longest side of a right-angled triangle and is opposite the 90° angle. The other two sides of the triangle are each shorter than the hypotenuse and form the 90° angle with each other. These two sides are the perpendicular components of the original vector.

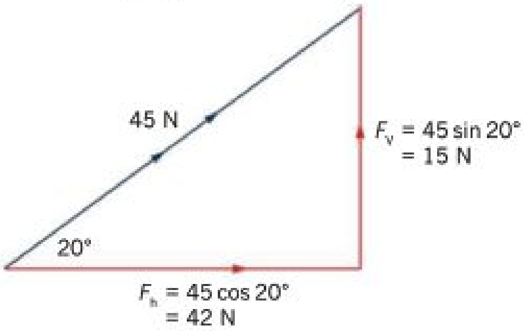


FIGURE 8.4.2 The perpendicular components (shown in red) of the original vector (shown in blue). The original vector is the hypotenuse of the triangle.

Geometric method of finding vector components

The geometric method of finding the perpendicular components of vectors is to construct a right-angled triangle using the original vector as the hypotenuse. This was illustrated in Figure 8.4.2. The magnitude and direction of the components are then determined using trigonometry. A good rule to remember is that no component of a vector can be larger than the vector itself. In a right-angled triangle, no side is longer than the hypotenuse. The original vector must be the hypotenuse and its components must be the other two sides of the triangle.

Figure 8.4.3 on page 280 shows a force vector of 50.0 N (drawn in black) acting on a box in a direction 30.0° up from horizontal to the right. The horizontal and vertical components of this force must be found in order to complete further calculations.

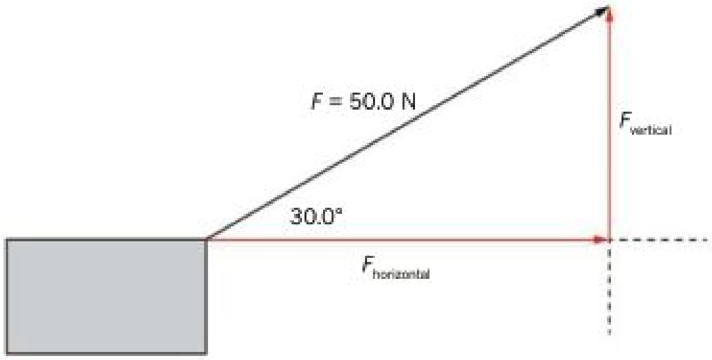


FIGURE 8.4.3 Finding the horizontal and vertical components of a force vector.

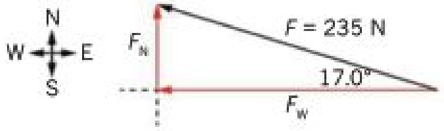
The horizontal component vector is drawn from the tail of the 50.0 N vector towards the right, with its head directly below the head of the original 50.0 N vector. The vertical component vector is drawn from the head of the horizontal component to the head of the original 50.0 N vector.

Using trigonometry, the horizontal component of the force is calculated to be 43.3 N horizontally to the right. The vertical component is calculated to be 25.0 N vertically upwards. The calculations are shown below:

$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$ adj = hyp $\cos\theta$ $F_h = (50.0)(\cos 30.0°)$ $= 43.3$ N horizontal to the right	$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$ opp = hyp $\sin\theta$ $F_V = (50.0)(\sin 30.0°)$ $= 25.0$ N vertically upwards

Worked example 8.4.1

CALCULATING THE PERPENDICULAR COMPONENTS OF A FORCE

<table>
<tr><td colspan="2">Use the direction conventions to determine the perpendicular components of a 235 N force acting on a bike at a direction of 17.0° north of west.</td></tr>
<tr><td>Thinking</td><td>Working</td></tr>
<tr><td>Draw F_W from the tail of the 235 N force along the horizontal direction, then draw F_N from the horizontal vector to the head of the 235 N force.</td><td>
</td></tr>
<tr><td>Calculate the west component of the force F_W using
$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$</td><td>$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$
adj = hyp $\cos\theta$
$F_W = (235)(\cos 17.0°)$
$= 224.7$ N west</td></tr>
<tr><td>Calculate the north component of the force F_N using
$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$</td><td>$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$
opp = hyp $\sin\theta$
$F_N = (235)(\sin 17.0°)$
$= 68.7$ N north</td></tr>
</table>

Worked example: Try yourself 8.4.1

CALCULATING THE PERPENDICULAR COMPONENTS OF A FORCE

Use the direction conventions to determine the perpendicular components of a 3540 N force acting on a trolley at a direction of 26.5° anticlockwise from the left direction.

8.4 Review

SUMMARY

- A vector can be resolved into two perpendicular component vectors.
- Perpendicular component vectors are at right angles to each other.
- Any component vectors must be smaller in magnitude than the original vector.
- The hypotenuse of a right-angled triangle is the longest side of the triangle and the other two sides are each smaller than the hypotenuse.
- A right-angled triangle vector diagram can be drawn with the original vector as the hypotenuse and the perpendicular components drawn from the tail of the original to the head of the original.
- The perpendicular components can be determined using trigonometry.

KEY QUESTIONS

1. Rayko applies a force of 462 N on the handle of a mower in a direction of 35.0° clockwise down from the right direction.
   - **a** What is the downwards force applied?
   - **b** What is the rightwards force applied?
2. A force of 25.9 N acts in the direction of S 40.0° E. Find the perpendicular components of the force.
3. A ferry is transporting students to Rottnest Island. At one point in the journey the ferry travels at 18.3 m s^{-1} N 75.6° W. Calculate its velocity in the northerly direction and in the westerly direction at that time.
4. Zehn walks 47.0 m in the direction of S 66.3° E across a hockey field. Calculate the change in Zehn's position down the field and across the field during that time.
5. A cargo ship has two tugs attached to it by ropes. One of the tugs is pulling directly north, while the other tug is pulling directly west. The pulling forces of the tugboats combine to produce a total force of 235 000 N in a direction of N 62.5° W. Calculate the force that each tug boat applies to the cargo ship.
6. Resolve the following forces into their perpendicular components around the north–south line. In part d, use the horizontal and vertical directions.
   - **a** 100 N S 60° E
   - **b** 60 N north
   - **c** 300 N 160°T
   - **d** 3×10^5 N 30° upwards from the horizontal.
7. What are the horizontal and vertical components of a 300 N force that is applied along a rope at 60° to the horizontal and used to drag an object across a yard?

8.5 Mass and weight

The difference between mass and weight is sometimes misunderstood because the terms are used interchangeably in the English language. In physics, however, the two terms have different meanings. For instance, weight is a vector and mass is a scalar. The difference between these two terms is explained in this section.

MASS OF A BODY

Mass is a scalar quantity. In scientific contexts, mass is measured in kilograms (kg). In earlier science courses, **mass** may have been defined as 'the amount of matter in an object'. Since the late 1700s, the kilogram has been defined in terms of an amount of a standard material. At first, 1 litre of water at 4°C was used to define the kilogram. More recently an international mass standard has been introduced. This is a 1 kg cylinder of platinum–iridium alloy that is kept in Paris (see Figure 8.5.1). Copies are made from this standard and sent around the world to calibrate balances.

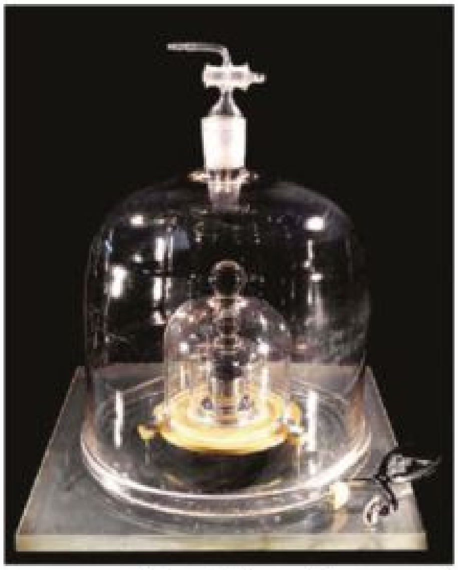


FIGURE 8.5.1 All mass in the world is compared to this small piece of platinum–iridium alloy held in a sealed vault in Paris.

PHYSICSFILE

Defining a kilogram

The kilogram is currently defined by the mass of a cylinder made of platinum–iridium alloy kept in Paris. It is the only SI unit that is defined in this way. A group of scientists from around the world, including Australian scientists, are currently looking at ways to redefine the kilogram in terms of a physical property that is unchanging and can be reproduced in laboratories. The Avogadro Project has been running for several years with the aim that soon there will be a scientific definition of the kilogram.

FIGURE 8.5.2 CSIRO scientists are working on a replacement for the standard kilogram.

To gain a better understanding of mass through the effect of a force on a massive body, think about a mass resting on a frictionless surface. If a force, *F*, is applied to the mass, *m*, in the horizontal direction, an acceleration, *a*, is produced as shown in Figure 8.5.3. That is, when an object experiences a force (push or pull) it will begin to move and while the force is acting, its velocity will change. When the velocity of an object changes, the object is said to be accelerating.

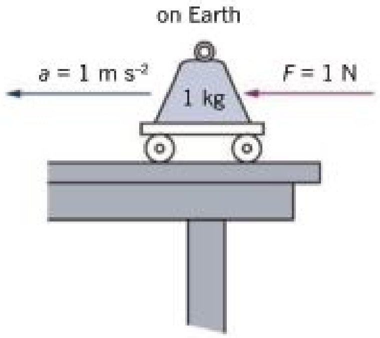


FIGURE 8.5.3 A force applied to an object can cause it to accelerate.

The more mass an object has, the greater the force required to make it accelerate. If the same force is applied to two different masses, the smaller mass will accelerate more than the greater mass. For this reason, mass can be seen as the property of a body that resists the change in motion caused by a force.

If the above experiment is repeated on the Moon with the same horizontal force acting on the body on a frictionless surface, the same acceleration will result, as shown in Figure 8.5.4. This is because the mass of the body remains the same on the Earth and on the Moon. Mass is a property of the body and it is not affected by its environment.

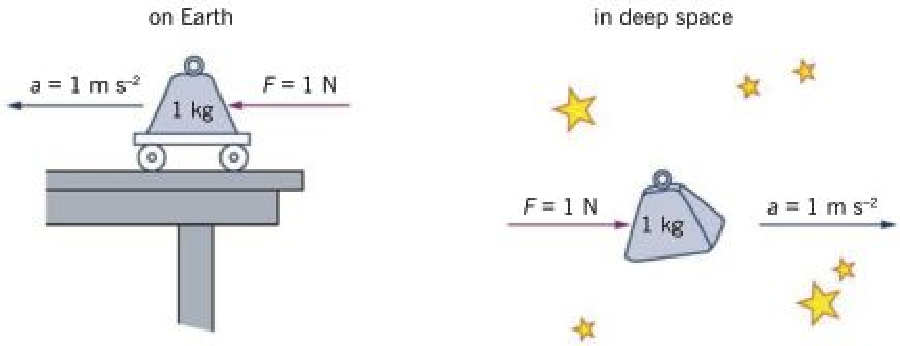


FIGURE 8.5.4 Mass is a property of the object and not of its surroundings.

GRAVITATIONAL FORCE

In the late 1500s, Galileo was able to show that all objects which are dropped near the surface of the Earth accelerate at the same rate, *g*, towards the centre of the Earth. The force that produces this acceleration is the force due to gravity. The force due to gravity is an attractive force that exists between all masses. In other words, it is a 'pulling' force that exists between everything that has a mass. It is one of the fundamental forces that acts over a distance, which means that the two masses do not need to be in contact in order for the force to exist.

PHYSICSFILE

The force of gravity between the Earth and the Moon

The Moon stays in orbit around the Earth due to the attractive force of gravity that acts between the two bodies. The force required to maintain the Moon in an orbit of the Earth is very large. If it could be replaced with a steel rod that connects the Earth and the Moon, the rod would have to be over 700 km in diameter.

Gravitational forces result from a mass creating a gravitational field that spreads throughout the space around the mass. Any other mass that is within this field will experience a force towards the mass creating the field. The object that is in the gravitational field also has mass, so it too has a gravitational field around it that attracts the original mass with an equal and opposite force.

The gravitational field extends through space in an inverse squared relationship. This means that if you double the distance from the mass creating the field, then the force will be one-quarter the size.

Here on Earth, you are strongly affected by the Earth's gravitational field (as well as by fields from the Sun, the Moon and other objects in the solar system). Even if you were not on Earth, you could still measure the effect of the Earth's gravitational field. There is no place in the universe where the Earth's gravitational field will not reach. At the 'edge' of the universe it will be very small, but it can be calculated. The closer you or any mass is located to the Earth, the larger the gravitational force of attraction towards the Earth. At a height above the Earth's surface that is equal to the radius of the Earth, the force due to gravity on a mass will be one-quarter of that at the Earth's surface. At two Earth radii above the Earth's surface, the gravitational force will be one-ninth of the force experienced on the Earth's surface.

WEIGHT FORCE

In physics, the force on a body due to gravity is called the **weight** of a body, F_g or just *W*. Weight is a force, therefore it is a vector quantity. Like other forces, it is measured in newtons (N).

Figure 8.5.5 on page 284 shows a bin falling through the air. As it falls, it accelerates downwards due to the Earth's gravitational field strength, *g*, which near the surface of the Earth is 9.81 N kg^{-1} down.

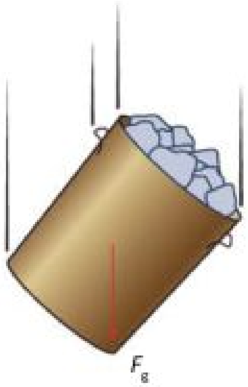


FIGURE 8.5.5 When the bin is in mid-air, there is an unbalanced force due to gravity acting on it so it accelerates towards the ground. This is the object's weight force. The vector representing weight is drawn from the centre of mass of the object and points downwards.

As the weight of the bin is a vector, it can be represented with an arrow. An arrow representing weight is drawn downwards (towards the centre of the Earth) with its tail beginning at the object's centre of mass. Centre of mass is described in detail in Chapters 9 and 11. Stated simply, an object's centre of mass is the point where its mass can be considered to be 'concentrated'. In an object of uniform density, there is as much mass above the centre of mass as there is below it, as much mass to the left as there is to the right, and as much mass in front as there is behind it.

The *weight* of a body F_g (in N) is defined as the force of attraction on a body due to gravity and is calculated using the equation:

$F_g = mg$

where F_g is the force of gravity acting at the centre of mass of a body (in N)

m is the mass of the body (in kg)

g is the gravitational field strength (in N kg^{-1}, which is 9.8 N kg^{-1} near the surface of the Earth).

Your mass and weight on the Moon

If you were ever lucky enough to travel to the Moon, you would notice that over the duration of the trip, the amount of matter that makes up your body would not change. There would still be as much of you present when you arrived on the Moon as there was when you left the Earth. Your mass wouldn't have changed because mass is a property of the matter and is not affected by its environment. However, you would notice when you stood on the floor of the Moon base that you were not pulled down as hard on the ground. In other words, your weight force would be much less on the Moon than it was on the Earth.

The Moon has a much smaller mass than the Earth. The Moon's mass is 7.348×10^{22} kg and the Earth's mass is 5.97×10^{24} kg. This means that the Moon is about 81 times smaller than the Earth. This smaller mass means that the Moon's gravitational field is much weaker than the Earth's gravitational field and therefore its force due to gravity on any object is much less. As the force is less, the acceleration that an object experiences on the Moon will be less than here on Earth. Consider Figure 8.5.6, in which a 5 kg pumpkin is falling on the Earth and then on the Moon.

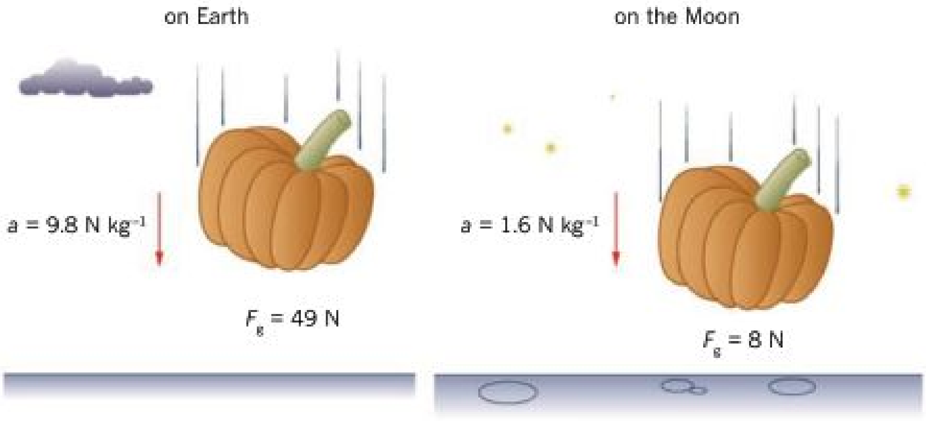


FIGURE 8.5.6 A 5 kg pumpkin falling on the Earth and on the Moon. The mass of the pumpkin is 5 kg no matter where it is, but the weight of the pumpkin is different.

The pumpkin has a much smaller weight force on the Moon than on the Earth due to the much smaller gravitational force.

8.5 Review

SUMMARY

- The standard mass is a 1 kg platinum–iridium cylinder, against which all other masses are compared.
- The mass of an object relates to its ability to resist changes in motion.
- Mass is a scalar quantity and is a property of a body that is not influenced by external environmental factors.
- Gravity is a force of attraction between masses that extends throughout space.
- Weight is a force due to gravity. As it is a force, it is also a vector and requires a magnitude and direction.
- Mass is measured in kilograms and weight is measured in newtons.
- The mass of an object will not change as the object goes from the Earth to the Moon, but the weight will change due to the decreased gravitational force on the Moon compared to the Earth.
- The centre of mass indicates the position at which the entire mass of a body is considered to be concentrated. At this point, all external forces are applied.

KEY QUESTIONS

1. An object is placed in a spaceship and launched into space. As the object leaves the surface of the Earth, it has a mass of 50 kg. What will be the object's mass if it lands on Pluto?
2. Two students have a conversation. One of the students states that her weight is 60 kg. What is wrong with this statement?
3. Mary's mass is 75 kg. What is her weight on Earth if g is 9.8 N kg^{-1}?
4. A desk chair has a weight of 34.3 N on the surface of the Earth. Determine the mass of the chair. Use $g = 9.8$ N kg^{-1}.
5. Determine the weight of the same chair in the previous question when it is on the surface of the Moon, where $g = 1.6$ N kg^{-1}.
6. On the surface of the Earth, a geological hammer has a mass of 1.5 kg. Determine its mass and weight on Mars, where $g = 3.6$ N kg^{-1}.
7. Would your weight be greater on Earth or on the Moon? Explain your answer.

Chapter review

08

KEY TERMS

collinear
component
dimension
direction conventions
magnitude
mass
resultant
scalar
units
vector
vector diagram
weight

1 Select the scalar quantities in the list below. (There may be more than one answer.)

A force
B time
C acceleration
D mass

2 Select the vector quantities in the list below. (There may be more than one answer.)

A displacement
B distance
C volume
D velocity

3 A basketballer applies a force with his hand to bounce the ball. Describe how a vector can be drawn to represent this situation.

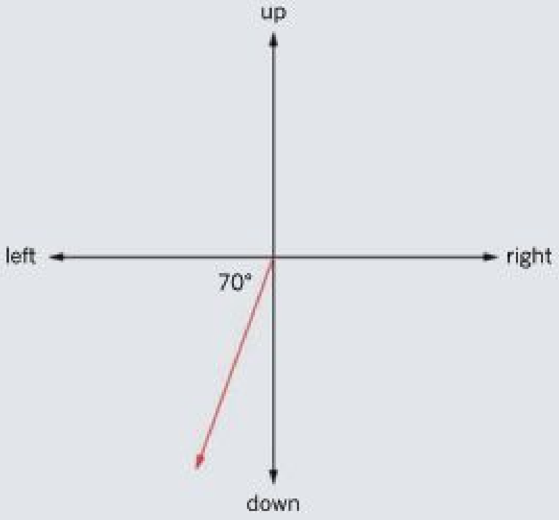
4 Vector arrow A is drawn twice the length of vector arrow B. What does this mean?

5 A car travels 15 m s^{-1} north and another travels 20 m s^{-1} south. Why is a sign convention often used to describe vectors like these?

6 When finding the change in velocity between an initial velocity of 34.0 m s^{-1} south and a final velocity of 12.5 m s^{-1} east, which two vectors need to be added together?

7 If the vector 20 N forwards is written as –20 N, how would you write a vector representing 80 N backwards?

8 Describe the following vector direction.



9 Add the following force vectors using a number line: 3 N left, 2 N right, 6 N right. Then also draw and describe the resultant force vector.

10 Determine the resultant vector of the following combination: 45.0 m forwards, 70.5 m backwards, 34.5 m forwards, 30.0 m backwards.

11 Find the vector which results from the addition of 36 m south and 55 m west.

12 Add the following vectors: 481 N north and 655 N east. Give answers to three significant figures.

13 Determine the change in velocity of a bird that changes from flying 3 m s^{-1} to the right to flying 3 m s^{-1} to the left.

14 A car makes a turn, changing its velocity from 13.0 m s^{-1} south to 18.7 m s^{-1} west. Calculate the change in the velocity vector, Δv, of the car, to three significant figures.

15 Bill hits a cricket ball so that it changes its velocity from 38.8 m s^{-1} east to 55.5 m s^{-1} north. Calculate the change in the velocity vector, to three significant figures.

16 A force of 45.5 N acts in the direction of S 60.0° E. Find the eastern and southern components of this force. Give your answers to three significant figures.

17 A pumpkin has a mass of 10 kg on Earth. What is its weight on Earth?

18 A skateboard has a weight of 20.6 N on Earth. What is its mass?

19 **a** What is the mass of an 85 kg astronaut on the surface of Earth where g is 9.8 N kg^{-1}?

b What is the mass of an 85 kg astronaut on the surface of the Moon where g is 1.6 N kg^{-1}?

c What is the weight of an 85 kg astronaut on the surface of Mars where g is 3.6 N kg^{-1}?

20 Given the figures in question 19, order the weight of a 1 kg object from greatest weight to least weight when it is on the Moon, on Mars and on Earth.

CHAPTER

09 Linear motion

Motion, from the simple to the complex, is a fundamental part of everyday life. The motion of a gymnast performing a floor routine is a complex form of motion. An Olympic snowboarder competing in a half-pipe event also exhibits a complex form of motion. Simpler examples include a skier travelling in a straight line down a ski run, a train pulling into a station and a swimmer completing a lap of a pool.

Key knowledge

By the end of this chapter you will have studied the simplest form of motion—straight-line motion—and will be able to:

- analyse graphically, numerically and algebraically straight-line motion under constant acceleration:
  $v = u + at$, $v^2 = u^2 + 2as$, $s = \frac{1}{2}(u + v)t$, $s = ut + \frac{1}{2}at^2$, $s = vt - \frac{1}{2}at^2$
- graphically analyse non-uniform motion in a straight line.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

9.1 Displacement, speed and velocity

In order to analyse and communicate ideas about motion, it is important to understand the terms used to describe motion—even in its simplest form. In this section you will learn about some of the terms used to describe straight-line motion, such as position, distance, displacement, speed and velocity.

CENTRE OF MASS

When analysing motion, things are often more complicated than they first appear. As a freestyle swimmer travels at a constant speed of 2 m s^{-1}, their head and torso move forwards at this speed. The motion of their arms, however, is more complex. At times their arms move forwards through the air faster than 2 m s^{-1}, and at other times they are actually moving backwards through the water.

It is beyond the scope of this course to analyse such a complex motion. However, the motion of the swimmer can be simplified by treating the swimmer as a simple object located at a single point called the **centre of mass** or centre of gravity. The centre of mass is the balance point of an object. For a person, the centre of mass is located near the waist. The centres of mass of some everyday objects are shown in Figure 9.1.1. The concept of centre of mass and centre of gravity is discussed in more detail in Section 11.3 on page 394.

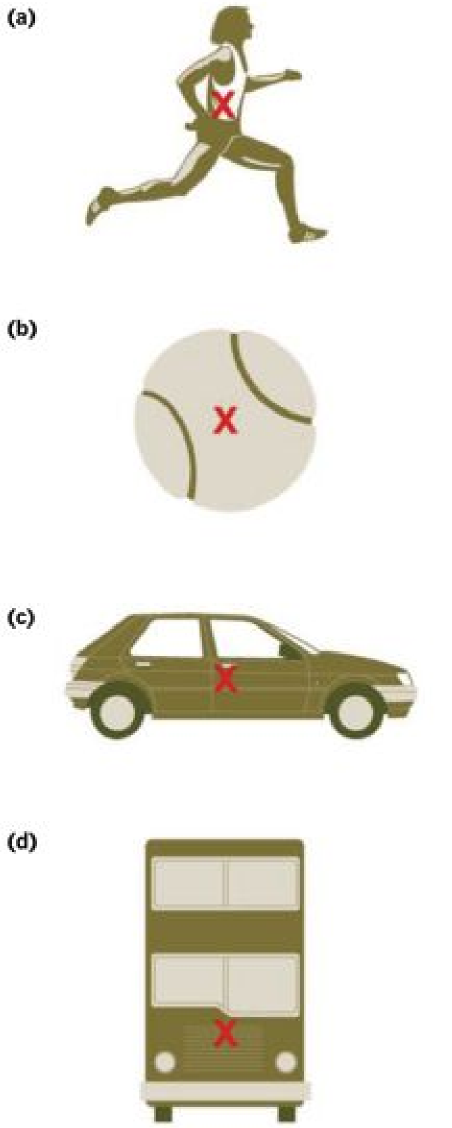


FIGURE 9.1.1 The centre of mass of each object is indicated by a cross.

POSITION, DISTANCE AND DISPLACEMENT

Position

One important term to understand when analysing straight line motion is **position**.

- Position describes the location of an object at a certain point in time with respect to the origin.
- Position is a vector quantity and therefore requires a direction.

Consider a swimmer, Sophie, doing laps in a 50 m pool, as shown in Figure 9.1.2. To simplify her motion, Sophie is treated as a simple point object. The pool can be treated as a one-dimensional number line, with the starting block as the origin. The direction to the right of the starting block is taken to be positive.

Sophie's position as she is warming up behind the starting block in Figure 9.1.2(a) is –10 m. The negative sign indicates the direction from the origin, i.e. to the left. Her position could also be given as 10 m to left of the starting block.

At the starting block (Figure 9.1.2(b)), Sophie's position is 0 m, then after swimming half a length of the pool she is +25 m or 25 m to the right of the origin (Figure 9.1.2(c)).

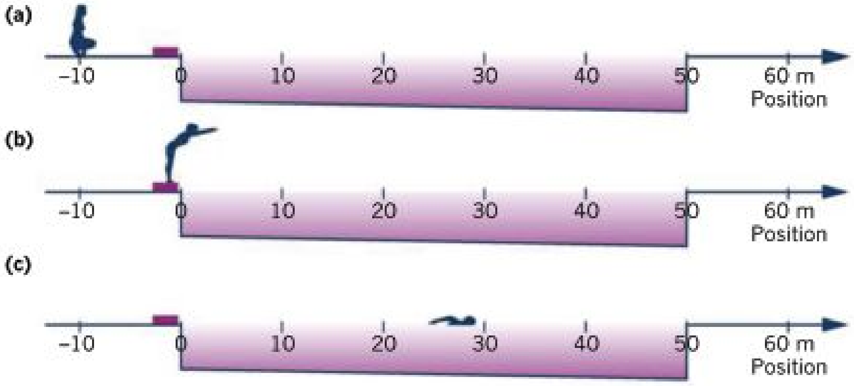


FIGURE 9.1.2 The position of the swimmer is given with reference to the starting block. (a) While warming up, Sophie is at –10 m. (b) When she is on the starting block, her position is zero. (c) After swimming for a short time, she is at a position of +25 m.

Distance travelled

Position describes where an object is at a certain point in time. However, **distance travelled** is how far a body travels during a journey.

- Distance travelled, *d*, describes the length of the path covered during an object's entire journey.
- Distance travelled is a scalar quantity and is measured in metres (m).

For example, if Sophie completes three lengths of the pool, the distance travelled during her swim will be 50 + 50 + 50 = 150 m.

The distance travelled is not affected by the direction of the motion. That is, the distance travelled by an object always increases as it moves, regardless of its direction. The tripmeter or odometer of a car or bike measures distance travelled.

Displacement

Displacement, *s*, is defined as the *change in position* of an object. Displacement takes into account only where the motion starts and finishes. The route taken between these points has no effect on displacement. The sign of the displacement indicates the direction in which the position has changed from the start to the end.

- Displacement is the change in position of an object in a given direction.
- Displacement s = final position – initial position.
- Displacement is a vector quantity and is measured in metres (m).

Consider the example of Sophie completing one length of the pool. During her swim, the distance travelled is 50 m. Her final position is +50 m and her initial position is 0 m. Her displacement is:

s = final position – initial position
= 50 – 0
= +50 m or 50 m in a positive direction

Notice that **magnitude**, units and direction are required for a vector quantity. The distance will be equal to the magnitude of displacement only if the body is moving in a straight line and does not change direction. If Sophie swims two lengths, her distance travelled will be 100 m: 50 m out and 50 m back. However, her displacement during this swim will be:

s = final position – initial position
= 0 – 0
= 0 m

Even though Sophie has swum 100 m, her displacement is zero because the initial and final positions are the same.

The above formula for displacement is useful if you already know the initial and final positions of a body's motion. An alternative method to determine total displacement, if you know the displacement of each section of the motion, is to add up the individual displacements for each section of motion.

total displacement = sum of individual displacements

It is important to remember that displacement is a vector and so, when adding displacements, you must obey the rules of vector addition (discussed in Chapter 8).

In the example above, in which Sophie completed two laps, overall displacement could have been calculated by adding the displacement of each lap:

s = sum of displacements for each lap
= 50 m in the positive direction + 50 m in the negative direction
= 50 + (–50)
= 0 m

SPEED AND VELOCITY

For thousands of years, humans have tried to travel at ever greater speeds. This desire has contributed to the development of all sorts of competitive activities, as well as major advances in engineering and design. World records for some of these pursuits are given in Table 9.1.1.

Activity	World record speed (m s^{-1})	World record speed (km h^{-1})
luge	43	140
train	161	575
tennis serve	73.1	263
waterskiing (barefoot)	68.3	246
cricket delivery	44.7	161
racehorse	19.7	71

TABLE 9.1.1 World record speeds for a variety of sports or modes of transport.

PHYSICS IN ACTION

Timing and false starts in athletics

Until 1964, all timing of events at the Olympic Games was recorded by handheld stopwatches (Figure 9.1.3). The reaction times of the judges meant an uncertainty of 0.2 s for any measurement. An electronic quartz timing system introduced in 1964 improved accuracy to 0.01 s, but in close finishes the judges still had to wait for a photograph of the finish before they could announce the places.

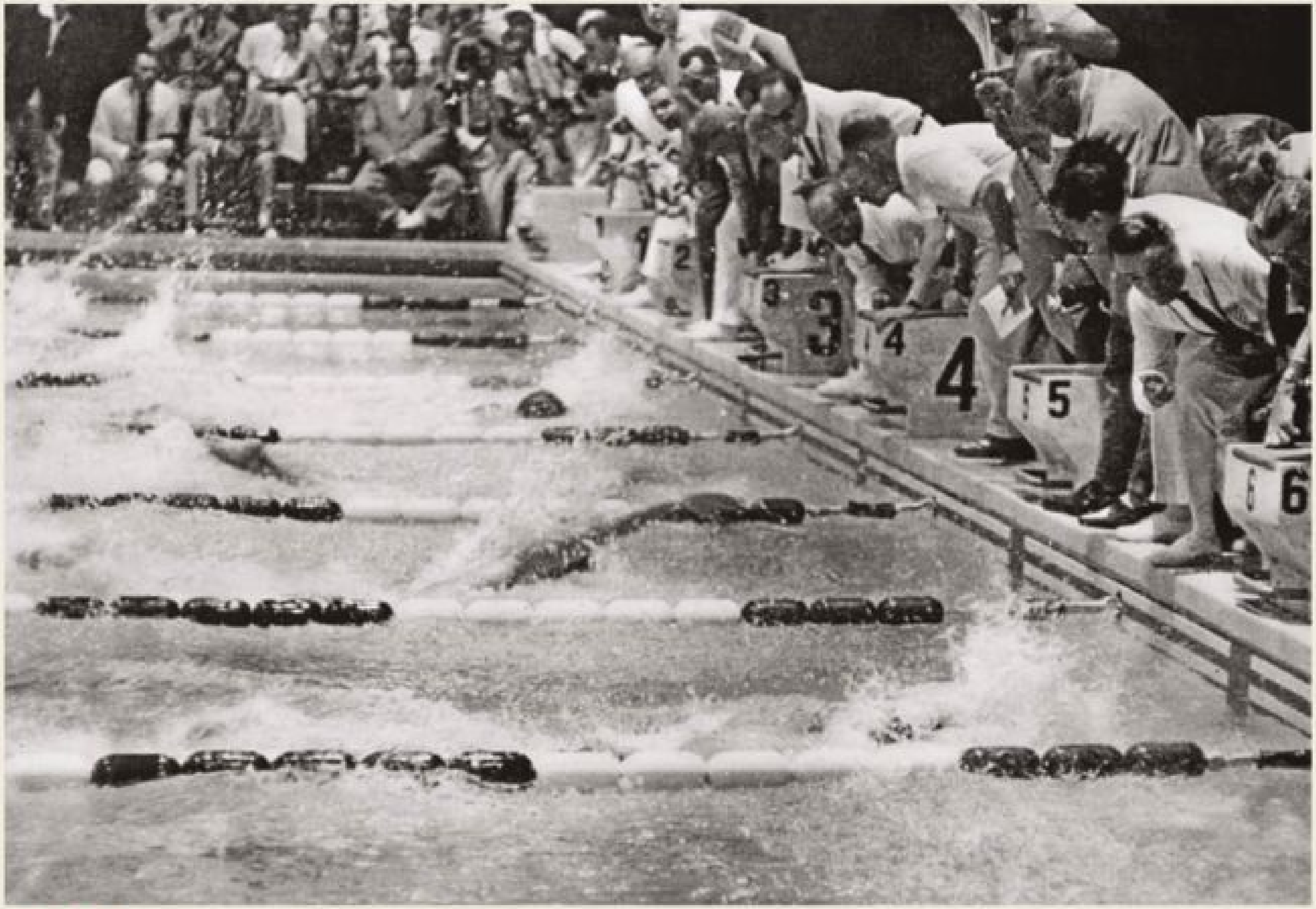


FIGURE 9.1.3 Using stopwatches to time a race in Rome in 1960.

The current timing system used in athletics is a vertical line-scanning video system (VLSV). Introduced in 1991, this electronic timing system is completely automatic. The starting pistol triggers a computer to begin timing. At the finish line, a high-speed video camera records the image of each athlete and indicates the time at which each one crosses the line. This system enables the times of all athletes in the race to be precisely measured to one-thousandth of a second.

FIGURE 9.1.4 Starting blocks are fitted with pressure sensors to detect false starts.

Another feature of this system is that it indicates when a runner 'breaks' at the start of the race. Each starting block is connected by electronic cables to the timing computer and a pressure sensor indicates if a runner has left the blocks early (Figure 9.1.4). A reaction time of 0.10 s has been incorporated into the system since 2002. This ensures that a runner has not anticipated the pistol. It also means that a runner can still commit a false start even if their start was after the pistol. A start that is less than 0.10 s after the pistol is registered as false.

The most controversial false start of recent times occurred at the athletics World Championships in 2003. It was a quarter-final heat of the men's 100 metres. US runner Jon Drummond was in lane 4 and Asafa Powell of Jamaica was in lane 5. Australia's Patrick Johnson was in lane 6. There had already been a false start in this heat and, since 2002, the rule for false starts in athletics events had been that after one false start, the next athlete to break is disqualified. The athletes went under starter's orders a second time and again there was a false start. The officials examined the computer read-out from the pressure pads on the blocks and determined that both Drummond and Powell were to be disqualified. Asafa Powell immediately left the track. Jon Drummond protested his innocence and proceeded to sit and lie down on the track for the next 20 minutes (Figure 9.1.5). He was widely criticised for his actions, but an analysis of the pressure pad readings revealed that he may have been a little unlucky.

FIGURE 9.1.5 Athlete Jon Drummond protested his false start by lying down on the track.

Speed and **velocity** are both quantities that give an indication of how quickly the position of an object is changing. Both terms are in common use and are often assumed to have the same meaning. In physics, however, these two terms have different definitions.

- Speed is defined in terms of the rate at which the distance is travelled. Like distance, speed is a scalar. A direction is not required when describing the speed of an object.
- Velocity is defined in terms of the rate at which displacement changes, and so is a vector quantity. A direction should always be given with a velocity.
- The SI unit for speed and velocity is metres per second (m s^{-1}), but kilometres per hour (km h^{-1}) is also commonly used.

Instantaneous speed and velocity

Instantaneous speed and instantaneous velocity give a measure of how fast something is moving at a particular point in time. The speedometer on a car or bike indicates instantaneous speed.

If a speeding car is travelling north and is detected on a police radar gun at 150 km h^{-1}, it indicates that this car's instantaneous speed is 150 km h^{-1}, while its instantaneous velocity is 150 km h^{-1} north. Notice that the instantaneous speed is equal to the magnitude of the instantaneous velocity. This is always the case for instantaneous speed and velocity.

Average speed and velocity

Average speed and *average velocity* both give an indication of how fast an object is moving over a time interval.

$$\text{average speed } v_{av} = \frac{\text{distance travelled}}{\text{time taken}} = \frac{d}{\Delta t}$$

$$\text{average velocity } v_{av} = \frac{\text{displacement}}{\text{time taken}} = \frac{s}{\Delta t}$$

Average speed is equal to instantaneous speed only when a body is moving in uniform motion (that is, if it moves at a constant speed).

The average speed of a car that takes 30 minutes to travel 20 km from St Kilda to Dandenong is 40 km h^{-1}. However, this does not mean that the car travelled the whole distance at this speed. In fact it is more likely that the car was moving at 60 km h^{-1} for a significant amount of time, while for some of the time the car would not be moving at all.

A direction (such as north, south, up, down, left, right, positive, negative) must be given when describing a velocity. The direction will always be the same as that of the displacement. Similar to the relationship between distance and displacement, average speed will be equal to the magnitude of average velocity only if the body is moving in a straight line and does not change direction.

For example, in a race around a circular track like the velodrome shown in Figure 9.1.6, regardless of the average speed for a complete lap the magnitude of the average velocity will be zero, because the displacement is zero.

FIGURE 9.1.6 Anna Meares won the UCI world championship in 2013. She rode 500 m in a world record time of 32.836 s. Her average speed was 55.6 km h^{-1} but her average velocity was zero.

EXTENSION

How police measure the speeds of cars

Road accidents cause the deaths of about 1200 people in Australia each year and many times this number are seriously injured. Numerous steps have been taken to reduce the number of road fatalities. Some of these include random alcohol and drug testing, speed cameras, mandatory wearing of bicycle helmets and the zero blood alcohol level for probationary drivers.

One of the main causes of road trauma is speeding. In their efforts to combat speeding motorists, police employ a variety of speed-measuring devices. One such device is shown in Figure 9.1.7.

FIGURE 9.1.7 Speed cameras on poles.

Speed camera radar

Camera radar units are usually placed in parked, unmarked vehicles. These units emit a radar signal frequency of 24.15 GHz (2.415×10^{10} Hz). The radar antenna has a parabolic reflector that enables the unit to produce a directional radar beam that is 5° wide, allowing individual vehicles to be targeted. The radar range and field of vision for a camera is shown in Figure 9.1.8. The radar signal allows speeds to be determined by the Doppler principle, where the reflected radar signal from an approaching vehicle has a higher frequency than the original signal. Similarly, the reflected signal from a receding vehicle has a lower frequency. This change in frequency or 'Doppler shift' is processed by the unit and gives a measurement of the instantaneous speed of the target vehicle.

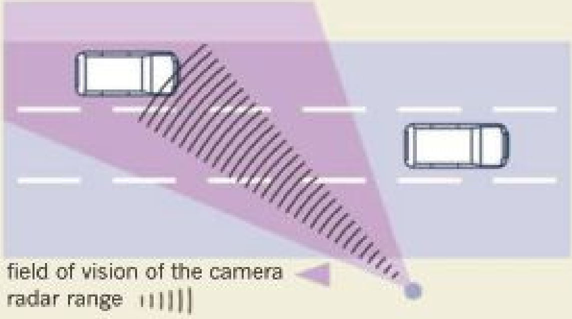


FIGURE 9.1.8 Diagram showing the visual range of a speed camera.

Camera radar units are capable of targeting a single vehicle up to 1.2 km away. In traffic, the units can distinguish between individual cars and take two photographs per second. The photographs and infringement notices are mailed to the offending motorists.

Laser speed guns

Speed guns are used by police to obtain an instant measure of the speed of an approaching or receding vehicle. The unit is usually handheld and is aimed directly at a vehicle using a target sight. It emits a pulse of infrared radiation frequency of 331 THz (3.31×10^{14} Hz). As with camera radar units, the speed is determined by the Doppler shift produced by the target vehicle. The infrared pulse is very narrow and directional, just 0.17° wide. This allows vehicles to be targeted with great precision. Handheld units can be used at distances up to 800 m. If the vehicle's speed registers over the limit, police are likely to pull the driver over.

Fixed speed cameras

Fixed speed cameras obtain their readings by using a system of three strips with piezoelectric sensors in them across the road (see Figure 9.1.9). The strips respond to the pressure as the car drives over them and create an electrical pulse that is detected. By knowing the distance between the strips and measuring the time that the car takes to travel across them, the speed of the car can be determined. This is actually measuring the average speed of the car, but by placing the strips close together the average speed gives a very good approximation of the instantaneous speed.

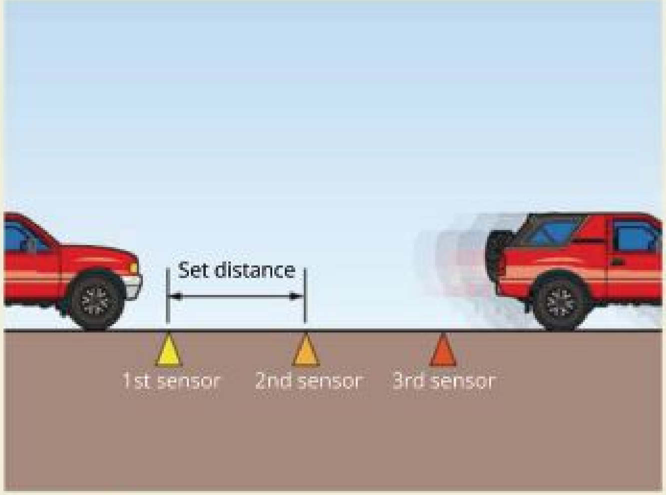


FIGURE 9.1.9 Fixed speed cameras record the speed of a car twice by measuring the time the car takes to travel over a series of three sensor strips embedded in the roadway.

PHYSICSFILE

Reaction time

Drivers are often distracted by loud music or phone calls. These distractions result in many accidents and deaths on the road. If cars are travelling at high speeds, they will travel a large distance in the time that the driver takes just to apply the brakes. A short reaction time is very important for all road users. This is easy to understand given the relationship between speed, distance and time.

distance travelled = $v \times t$

Converting km h^{-1} to m s^{-1}

You should be familiar with 100 km h^{-1} as it is the speed limit for most freeways and country roads in Australia. Cars that maintain this speed would travel 100 km in 1 hour. Since there are 1000 metres in 1 kilometre and 3600 seconds in 1 hour (60 s × 60 min), this is the same as travelling 100 000 m in 3600 s.

$$\begin{aligned}100\ \text{km h}^{-1} &= 100 \times 1000\ \text{m h}^{-1}\\ &= 100\,000\ \text{m h}^{-1}\\ &= \tfrac{100\,000}{3600}\ \text{m s}^{-1}\\ &= 27.8\ \text{m s}^{-1}\end{aligned}$$

So km h^{-1} can be converted to m s^{-1} by multiplying by $\frac{1000}{3600}$ (or dividing by 3.6).

Converting m s^{-1} to km h^{-1}

A champion Olympic sprinter can run at an average speed of close to 10 m s^{-1}. Each second, the athlete will travel approximately 10 metres. At this rate, in 1 hour the athlete would travel 10 × 3600 = 36 000 m = 36 km.

$$\begin{aligned}10\ \text{m s}^{-1} &= 10 \times 3600\ \text{m h}^{-1}\\ &= 36\,000\ \text{m h}^{-1}\\ &= \tfrac{36\,000}{1000}\ \text{km h}^{-1}\\ &= 36\ \text{km h}^{-1}\end{aligned}$$

So m s^{-1} can be converted to km h^{-1} by multiplying by $\frac{3600}{1000}$ or 3.6.

When converting a speed from one unit to another, it is important to think about the speeds to ensure that your answers make sense. The diagram in Figure 9.1.10 summarises the conversion between units for speed.

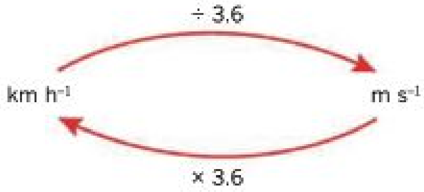


FIGURE 9.1.10 Rules for converting between m s^{-1} and km h^{-1}.

PHYSICS IN ACTION

Alternative units for speed and distance

Metres per second is the standard unit for measuring speed because it is derived from the standard unit for distance (metres) and the standard unit for time (seconds). However, alternative units are often used to better suit a certain application.

The speed of a boat is usually measured in knots, where 1 knot = 0.51 m s^{-1}. This unit originated in the nineteenth century, when the speed of sailboats would be measured by allowing a rope, with knots tied at regular intervals, to be dragged by the water through a sailor's hands. By counting the number of knots that passed through the sailor's hands, and measuring the time taken for this to happen, the average speed formula could be applied to estimate the speed of the boat.

The speed of very fast aeroplanes, such as the one in Figure 9.1.11, can be measured in Mach numbers. One Mach (referred to as Mach 1) is equal to the speed of sound, which is 340 m s^{-1}. Mach 2 is equal to 680 m s^{-1}, or twice the speed of sound.

The light-year is an alternative unit for measuring distance. The speed of light in a vacuum is nearly 300000 km s^{-1}. One light-year is the distance that light travels in one year. Because distances between objects in the universe are so large, astronomers use the light-year to measure distances. It takes over 4 years for light to travel from our nearest star (Alpha Centauri) to Earth. That means the distance from Earth to our nearest star is over 4 light-years. Light takes approximately 8.5 minutes to travel from the Sun to Earth, so it could be said that the Sun is 8.5 light-minutes away.

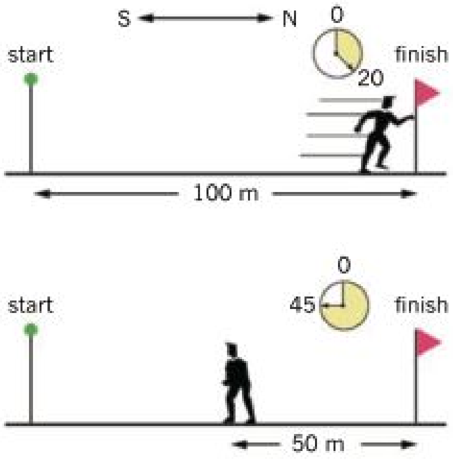
FIGURE 9.1.11 Modern fighter aeroplanes are able to fly at speeds above Mach 1.

Worked example 9.1.1

AVERAGE VELOCITY AND CONVERTING UNITS

Sam is an athlete performing a training routine by running back and forth along a straight stretch of running track. He jogs 100 m north in a time of 20 s, then turns and walks 50 m south in a further 25 s before stopping.

a What is Sam's average velocity in m s^{-1}?

Thinking	Working
Calculate the displacement. Remember total displacement is the sum of individual displacements. Sam's total journey consists of two displacements: 100 north and then 50 m south.	s = sum of displacements = 100 m north + 50 m south = 100 + (−50) = +50 m or 50 m north 
Work out the total time taken for the journey.	20 + 25 = 45 s
Substitute the values into the velocity equation.	Displacement is 50 m north. Time taken is 45 s. Average velocity $v_{av} = \frac{s}{\Delta t}$ $= \frac{50}{45}$ = 1.1 m s^{-1}
Velocity is a vector, so a direction must be given.	1.1 m s^{-1} north

b What is the magnitude of Sam's average velocity in km h^{-1}?

Thinking	Working
Convert from m s^{-1} to km h^{-1} by multiplying by 3.6.	v_{av} = 1.1 m s^{-1} = 1.1 × 3.6 = 4.0 km h^{-1} north
As the magnitude of velocity is needed, direction is not required in this answer.	Magnitude of v_{av} = 4.0 km h^{-1}

c What is Sam's average speed in m s^{-1}?	
Thinking	**Working**
Calculate the distance. Remember distance is the length of the path covered in the entire journey. The direction does not matter. Sam travels 100 m in one direction and then 50 m in the other direction	$d = 100 + 50$ $= 150$
Work out the total time taken for the journey.	20 + 25 = 45 s
Substitute the values into the speed equation.	Distance is 150 m. Time taken is 45 s. Average speed $v_{av} = \frac{d}{\Delta t}$ $= \frac{150}{45}$ $= 3.3$ m s^{-1}

d What is Sam's average speed in km h^{-1}?	
Thinking	**Working**
Convert from m s^{-1} to km h^{-1} by multiplying by 3.6.	Average speed $v_{av} = 3.3$ m s^{-1} $= 3.3 \times 3.6$ $= 12$ km h^{-1}

Worked example: Try yourself 9.1.1

AVERAGE VELOCITY AND CONVERTING UNITS

Sally is an athlete performing a training routine by running back and forth along a straight stretch of running track. She jogs 100 m west in a time of 20 s, then turns and walks 160 m east in a further 45 s before stopping.

a What is Sally's average velocity in m s^{-1}?

b What is the magnitude of Sally's average velocity in km h^{-1}?

c What is Sally's average speed in m s^{-1}?

d What is Sally's average speed in km h^{-1}?

PHYSICS IN ACTION

Breaking the speed limit

Over the past 100 years, advances in engineering and technology have led to the development of faster machines. Cars, planes and trains can now move people at speeds that were thought to be both unattainable and life-threatening a century ago.

The 1-mile land-speed record is 1220 km h^{-1} (339 m s^{-1}). This was set in 1997 in Nevada by Andy Green driving his jet-powered Thrust SSC.

The fastest combat jet is the MiG-25. In 1976 it reached a speed of 3800 km h^{-1} (1056 m s^{-1}), which is more than three times the speed of sound.

The fastest speed recorded by a train is 575 km h^{-1} (160 m s^{-1}) in 2007 by the French TGV *Atlantique*, although it does not reach this speed during normal operations.

In 2007, Markus Stoeckl of Austria set a new speed record for mountain biking. He reached a speed of 210 km h^{-1} racing down a ski slope in Chile. He is pictured in Figure 9.1.12. This record was broken by Eric Barone in 2015 with a speed of 223.3 km h^{-1}.

FIGURE 9.1.12 Markus Stoeckl setting a new speed record for mountain biking in 2007.

9.1 Review

SUMMARY

- Position defines the location of an object with respect to a defined origin.
- Distance travelled, d, tells us how far an object has actually travelled. Distance travelled is a scalar.
- Displacement, s, is a vector and is defined as the change in position of an object in a given direction: s = final position – initial position.
- The average speed of a body, v_{av}, is defined as the rate of change of distance and is a scalar quantity: average velocity $v_{av} = \frac{\text{distance travelled}}{\text{time taken}} = \frac{d}{\Delta t}$
- The average velocity of a body, v_{av}, is defined as the rate of change of displacement and is a vector quantity: average velocity $v_{av} = \frac{\text{displacement}}{\text{time taken}} = \frac{s}{\Delta t}$
- To convert from m s^{-1} to km h^{-1}, multiply by 3.6.
- To convert from km h^{-1} to m s^{-1}, divide by 3.6.
- The SI unit for both speed and velocity is metres per second (m s^{-1}).

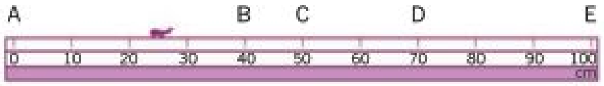
9.1 Review *continued*

KEY QUESTIONS

1 A girl swims ten lengths of a 25 m pool. Which one or more of the following statements correctly describes her distance travelled and displacement?

A Her distance travelled is zero.
B Her displacement is zero.
C Her distance travelled is 250 m.
D Her displacement is 250 m.

2 A somewhat confused ant is walking back and forth along a metre ruler, as show in the figure below. Taking the right as positive, determine both the size of the displacement and the distance travelled by the ant as it travels on the following paths.

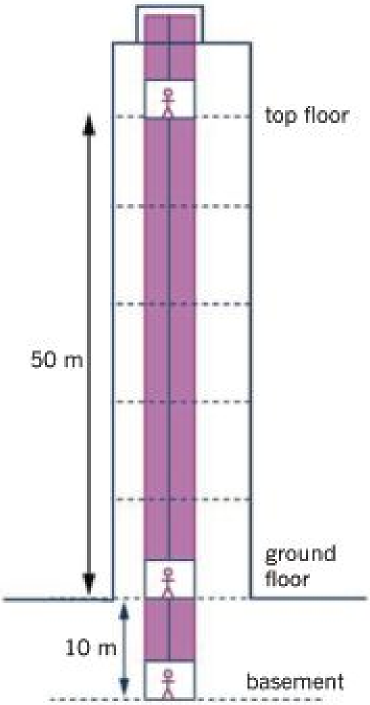


a A to B
b C to B
c C to D
d C to E and then to D

3 During a training ride, a cyclist rides 50 km north then 30 km south.

a What is the distance travelled by the cyclist during the ride?
b What is the displacement of the cyclist for this ride?

4 A lift in a city building, shown in the figure below, carries a passenger from the ground floor down to the basement, then up to the top floor.



a What is the displacement of the lift as it travels from the ground floor to the basement?
b What is the displacement of the lift as it travels from the basement to the top floor?
c What is the distance travelled by the lift during this entire trip?
d What is the displacement of the lift during this entire trip?

5 A car travelling at a constant speed was timed over 400 m and was found to cover the distance in 12 s.

a What was the car's average speed?
b The driver was distracted and his reaction time was 0.75 s before applying the brakes. How far did the car travel in this time?

6 A cyclist travels 25 km in 90 minutes.

a What is her average speed in km h^{-1}?
b What is her average speed in m s^{-1}?

7 Liam pushes his toy truck 5 m east, then stops it and pushes it 4 m west. The entire motion takes 10 seconds.

a What is the truck's average speed?
b What is the truck's average velocity?

8 Mihi rides her bicycle to school and travels 2.5 km south in 15 min.

a Calculate her average speed in kilometres per hour (km h^{-1}).
b What was her average velocity in metres per second (m s^{-1})?

9 An athlete in training for a marathon runs 10 km north along a straight road before realising that she has dropped her drink bottle. She turns around and runs back 3 km to find her bottle, then resumes running in the original direction. After running for 1.5 h, the athlete reaches 15 km from her starting position and stops.

a What is the distance travelled by the athlete during the run?
b What is the athlete's displacement during the run?
c What is the average speed of the athlete in km h^{-1}?
d What is the athlete's average velocity in km h^{-1}?

9.2 Acceleration

If you have been on a train as it pulled out of the station, you have experienced acceleration. If you have been in an aeroplane as it has taken off along a runway, you will have experienced a much greater acceleration. Astronauts and fighter pilots experience enormous accelerations that would make an untrained person lose consciousness. **Acceleration**, which is a measure of how quickly velocity changes, will be discussed in this section.

FINDING THE CHANGE IN VELOCITY AND SPEED

The velocity and speed of everyday objects are changing all the time. Examples of these are when a car moves away as the traffic lights turn green, when a tennis ball bounces or when you travel on a rollercoaster. If the initial and final velocity of an object are known, its change in velocity can be calculated.

To find the change, Δ, in any physical quantity, including speed and velocity, the initial value is taken away from the final value:

$$\Delta v = v - u$$

Change in speed is the final speed minus the initial speed:

$$\Delta v = v - u$$

where u is the initial speed (in m s^{-1})

v is the final speed (in m s^{-1})

Δv is the change in speed (in m s^{-1}).

Since speed is a scalar, direction is not required.

Change in velocity is the final velocity minus the initial velocity:

$$\Delta v = v - u$$

where u is the initial velocity (in m s^{-1})

v is the final velocity (in m s^{-1})

Δv is the change in velocity (in m s^{-1}).

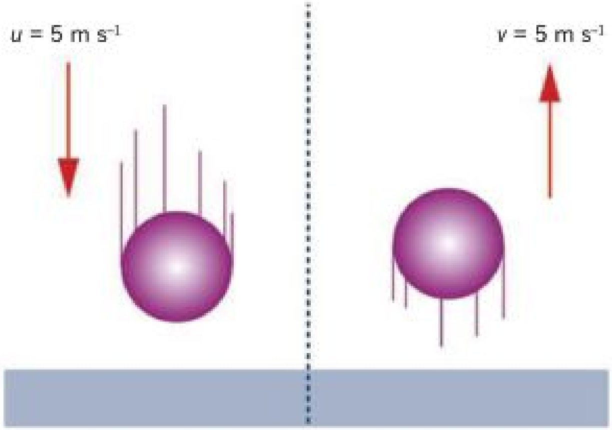
Since velocity is a vector, this should be done by performing a vector subtraction. As for all vectors, direction is required.

Vector subtraction was covered in detail in Section 8.3 on page 272.

Worked example 9.2.1

CHANGE IN SPEED AND VELOCITY PART 1

A golf ball is dropped onto a concrete floor and strikes the floor at 5.0 m s^{-1}. It then rebounds at 5.0 m s^{-1}.



a What is the change in speed of the ball?	
Thinking	**Working**
Find the values for the initial speed and the final speed of the ball.	u = 5.0 m s^{-1} v = 5.0 m s^{-1}
Substitute the values into the change in speed equation: $\Delta v = v - u$	$\Delta v = v - u$ $= (5.0) - (5.0)$ $= 0$ m s^{-1}

b What is the change in velocity of the ball?	
Thinking	**Working**
Velocity is a vector. Apply the sign convention to replace the directions.	u = 5.0 m s^{-1} down = −5.0 m s^{-1} v = 5.0 m s^{-1} up = +5.0 m s^{-1}
As this is a vector subtraction, reverse the direction of u to get $-u$.	u = −5.0 m s^{-1}, therefore $-u$ = +5.0 m s^{-1}
Substitute the values into the change in velocity equation: $\Delta v = v + (-u)$	$\Delta v = v + (-u)$ $= (+5.0) + (+5.0)$ $= +10$ m s^{-1}
Apply the sign convention to describe the direction.	Δv = 10.0 m s^{-1} up

Worked example: Try yourself 9.2.1

CHANGE IN SPEED AND VELOCITY PART 1

A golf ball is dropped onto a concrete floor and strikes the floor at 9.0 m s^{-1}. It then rebounds at 7.0 m s^{-1}.

a What is the change in speed of the ball?

b What is the change in velocity of the ball?

ACCELERATION

Consider the following information about the velocity of a car that starts from rest as shown in Figure 9.2.1.

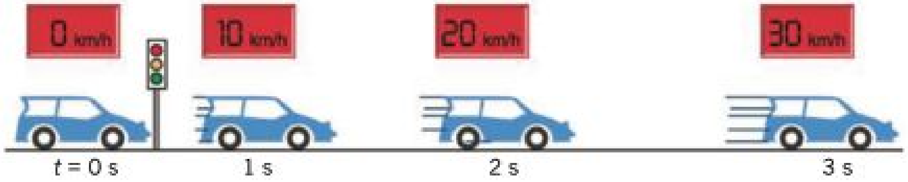


FIGURE 9.2.1 A car's acceleration as it increases in velocity from 0 km h^{-1} to 30 km h^{-1}.

The velocity of the car pictured above increases by 10 km h^{-1} each second. In other words, its velocity changes by +10 km h^{-1} per second. This is stated as an acceleration of +10 kilometres per hour per second or +10 km h^{-1} s^{-1}. More commonly in physics, velocity information is given in metres per second.

The athlete in Figure 9.2.2 takes 3 seconds to come to a stop at the end of a race.

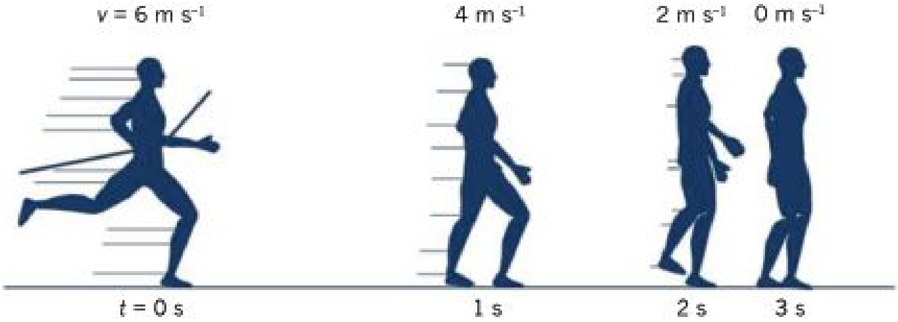


FIGURE 9.2.2 The velocity of the athlete changes by –2 m s^{-1} each second. The acceleration is –2 m s^{-2}.

The velocity of the athlete changes by –2 m s^{-1} each second, so the acceleration is –2 metres per second per second. This is usually expressed as –2 metres per second squared or –2 m s^{-2}.

A negative acceleration can mean that the object is slowing down in the direction of travel, as is the case with the athlete in Figure 9.2.2. A negative acceleration can also mean speeding up but in the opposite direction.

As acceleration is a vector quantity, vector diagrams can be used to calculate resultant accelerations of an object. Vector diagrams were covered in Chapter 8.

Average acceleration

As with speed and velocity, the average acceleration of an object can also be calculated.

Average acceleration, a_{av}, is the rate of change of velocity:

$$a_{av} = \frac{\text{change in velocity}}{\text{time taken}} = \frac{\Delta v}{\Delta t} = \frac{v - u}{\Delta t}$$

where v is the final velocity (in m s^{-1})

u is the initial velocity (in m s^{-1})

Δt is the time interval (s).

Worked example 9.2.2

CHANGE IN SPEED AND VELOCITY PART 2

A golf ball is dropped onto a concrete floor and strikes the floor at 5.0 m s^{-1}. It then rebounds at 5.0 m s^{-1}. The contact with the floor lasts for 25 ms.

What is the average acceleration of the ball during its contact with the floor?	
Thinking	**Working**
Note the values you will need in order to find the average acceleration (initial velocity, final velocity and time). Convert ms into s by dividing by 1000. (Note that Δv was calculated for this situation in the previous Worked example.)	$u = -5$ m s^{-1} $-u = 5$ m s^{-1} $v = 5$ m s^{-1} $\Delta v = 10$ m s^{-1} up $\Delta t = 25$ m s $= 0.025$ s
Substitute the values into the average acceleration equation.	$a_{av} = \frac{\text{change in velocity}}{\text{time taken}}$ $= \frac{\Delta v}{\Delta t}$ $= \frac{10}{0.025}$ $= 400$ m s^{-2}
Acceleration is a vector, so you must include a direction in your answer.	$a_{av} = 400$ m s^{-2} up

Worked example: Try yourself 9.2.2

CHANGE IN SPEED AND VELOCITY PART 2

A golf ball is dropped onto a concrete floor and strikes the floor at 9.0 m s^{-1}. It then rebounds at 7.0 m s^{-1}. The contact time with the floor is 35 ms.

What is the average acceleration of the ball during its contact with the floor?

PHYSICS IN ACTION

Human acceleration

In the 1950s, the United States Air Force used a rocket sled to determine the effect of extremely large accelerations on humans. One of these sleds is shown in Figure 9.2.3. The aim was to find out the greatest accelerations that humans could safely withstand in order to develop ejector seats for pilots.

The testing site consisted of an 800 m long railway track and a sled with nine rocket motors. One volunteer, Colonel John Stapp, was strapped into the sled and accelerated to speeds of over 1000 km h^{-1} in a very short time. Water scoops were used to stop the sled abruptly in just 0.35 s. This equates to a deceleration of greater than 400 m s^{-2}. The effects of these massive accelerations are evident on his face (Figure 9.2.4).

Colonel John Stapp was a human guinea pig who suffered a great deal of discomfort so that other pilots would benefit. Safer ejector seats and non-human crash test dummies were developed as a result of these experiments.

FIGURE 9.2.3 The rocket-powered sled used to test the effects of acceleration on humans.

FIGURE 9.2.4 Photos showing the distorted face of Colonel John Stapp.

9.2 Review

SUMMARY

- Change in speed is a scalar calculation:
  Δv = final speed – initial speed = $v - u$
- Change in velocity is a vector calculation:
  Δv = final velocity – initial velocity = $v - u$
- Acceleration is a vector. The average acceleration of a body, a_{av}, is defined as the rate of change of velocity:

$$a_{av} = \frac{\text{change in velocity}}{\text{time taken}} = \frac{\Delta v}{\Delta t} = \frac{v - u}{\Delta t}$$

- Acceleration is measured in metres per second per second (m s^{-2}).

KEY QUESTIONS

1 A radio-controlled car is travelling east at 10 km h^{-1}. It hits some sand and slows down to 3 km h^{-1} east. What is its change in speed?

2 A lump of Blu Tack is falling vertically at 5.0 m s^{-1} and as it hits the floor it stops dead. What is its change in velocity during the collision?

3 A ping pong ball is falling vertically at 5.0 m s^{-1}. As it hits the floor, it rebounds at 3.0 m s^{-1} up. What is its change in velocity during the bounce?

4 While playing soccer, Ashley is running north at 7.5 m s^{-1}. He slides along the ground and stops in 1.5 s. What is his average acceleration as he slides to a stop?

5 Olivia launches a model rocket vertically and it reaches a speed of 150 km h^{-1} after 3.5 s. What is the magnitude of its average acceleration in km h^{-1} s^{-1}?

6 A squash ball travelling east at 25 m s^{-1} strikes the front wall of the court and rebounds at 15 m s^{-1} west. The contact time between the wall and the ball is 0.050 s. Use vector diagrams, where appropriate, to help you with your calculations.

   a What is the change in speed of the ball?
   b What is the change in velocity of the ball?
   c What is the magnitude of the average acceleration of the ball during its contact with the wall?

7 A greyhound starts from rest and accelerates uniformly. Its velocity after 1.2 s is 8.0 m s^{-1} south.

   a What is the change in speed of the greyhound?
   b What is the change in velocity of the greyhound?
   c What is the magnitude of the acceleration of the greyhound?

9.3 Graphing position, velocity and acceleration over time

At times, even the motion of an object travelling in a straight line can be complicated. The object may travel forwards or backwards, speed up or slow down, or even stop. Where the motion remains in one dimension, the information can be presented in graphical form.

The main advantage of a graph compared with a table is that it allows the nature of the motion to be seen clearly. Information that is contained in a table is not as readily accessible or as easy to interpret as information presented graphically. This section examines position–time, velocity–time and acceleration–time graphs.

POSITION–TIME (x–t) GRAPHS

A position–time graph indicates the position, x, of an object at any time, t, for motion that occurs over an extended time interval. However, the graph can also provide additional information.

Consider Sophie, shown in Figure 9.3.1, swimming laps of a 50 m pool. Her position–time data are shown in Table 9.3.1. The starting point is treated as the origin for this motion.

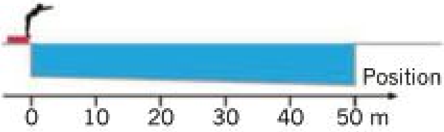


FIGURE 9.3.1 Swimmer standing at the end of a 50 m swimming pool.

Time (s)	0	5	10	15	20	25	30	35	40	45	50	55	60
Position (m)	0	10	20	30	40	50	50	50	45	40	35	30	25

TABLE 9.3.1 Positions and times of a swimmer completing 1.5 lengths of a pool.

Analysis of Table 9.3.1 reveals several features of Sophie's swim. For the first 25 s, she swims at a constant rate. Every 5 s she travels 10 m in a positive direction, i.e. her velocity is +2 m s^{-1}. Then, from 25 s to 35 s, her position does not change. She seems to be resting, as she is stationary for this 10 s interval. Finally, from 35 s to 60 s, she swims back towards the starting point, in a negative direction. On this return lap, she maintains a more leisurely rate of 5 m every 5 s, so her velocity is −1 m s^{-1}. However, Sophie does not complete this lap but ends 25 m from the start. This data is shown more conveniently on the position–time graph in Figure 9.3.2.

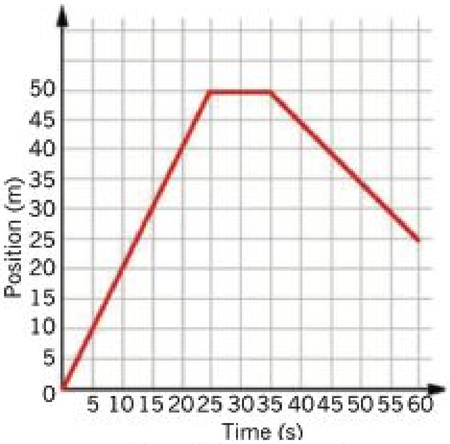


FIGURE 9.3.2 This position–time graph represents the motion of a swimmer travelling 50 m along a pool, then resting and swimming back towards the starting position. The swimmer finishes halfway along the pool.

The displacement, s, of the swimmer can be determined by comparing the initial and final positions. Her displacement between 20 s and 60 s is, for example:

s = final position − initial position

$= 25 - 40$

$= -15$ m

By further examining the graph above, it can be seen that during the first 25 s, the swimmer has a displacement of +50 m. Thus her average velocity is +2 m s^{-1}, i.e. 2 m s^{-1} to the right, during this time. This value can also be obtained by finding the gradient of this section of the graph.

Gradient of x–t graph = velocity

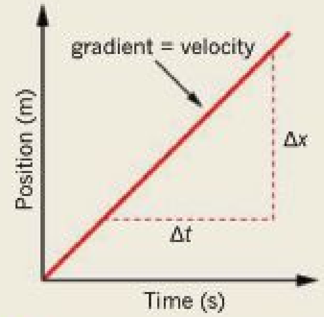


FIGURE 9.3.3 Position–time graph with gradient.

A positive velocity indicates that the object is moving in a positive direction and negative velocity indicates motion in a negative direction.

To confirm that the gradient of a position–time graph is a measure of velocity you can use **dimensional analysis**:

Gradient of x–t graph = $\frac{\text{rise}}{\text{run}} = \frac{\Delta x}{\Delta t}$

The units of this gradient will be metres per second (m s^{-1}) so gradient is a measure of velocity. Note that the rise in the graph is the change in position, which is the definition of displacement; that is, $\Delta x = s$.

Non-uniform velocity

For motion with uniform (constant) velocity, the position–time graph will be a straight line, but if the velocity is non-uniform the graph will be curved. If the position–time graph is curved, the instantaneous velocity will be the gradient of the tangent to the line at the point of interest; the average velocity will be the gradient of the chord between two points. This is illustrated in Figure 9.3.4.

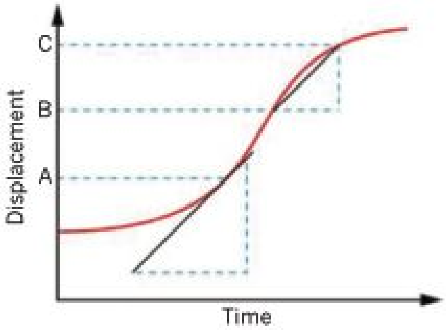
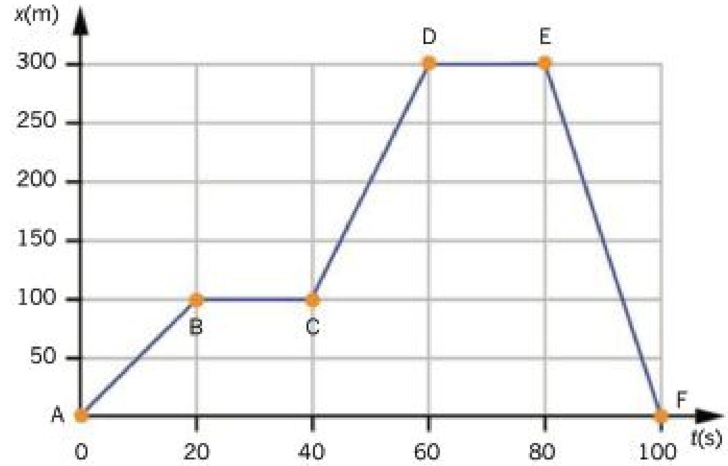


FIGURE 9.3.4 The instantaneous velocity at point A is the gradient of the tangent at that point. The average velocity between points B and C is the gradient of the chord between these points on the graph.

Worked example 9.3.1

ANALYSING A POSITION–TIME GRAPH

The motion of a cyclist is represented by the position–time graph below, with important features of the motion labelled A, B, C, D, E and F.



a What is the velocity of the cyclist between A and B?	
Thinking	**Working**
Determine the change in position (displacement) of the cyclist between A and B using: s = final position – initial position	At A, x = 0 m. At B, x = 100 m. $s = 100 - 0$ = +100 m or 100 m forwards (that is, away from the starting point)
Determine the time taken to travel from A to B.	$\Delta t = 20 - 0$ = 20 s
Calculate the gradient of the graph between A and B using: gradient of x–t graph = $\frac{\text{rise}}{\text{run}} = \frac{\Delta x}{\Delta t}$ Remember that $\Delta x = s$.	Gradient = $\frac{100}{20}$ = 5
State the velocity, using: gradient of x–t graph = velocity Velocity is a vector so direction must be given.	Since the gradient is 5, the velocity is +5 m s^{-1} or 5 m s^{-1} forwards.

b Describe the motion of the cyclist between B and C.

Thinking	**Working**
Interpret the shape of the graph between B and C.	The graph is flat between B and C, indicating that the cyclist's position is not changing for this time. So the cyclist is not moving. If the cyclist is not moving, the velocity is 0 m s^{-1}.
You may confirm the result by calculating the gradient of the graph between B and C using: gradient of x–t graph $= \frac{\text{rise}}{\text{run}} = \frac{\Delta x}{\Delta t}$ Remember that $\Delta x = s$.	Gradient $= \frac{0}{20}$ $= 0$

Worked example: Try yourself 9.3.1

ANALYSING A POSITION–TIME GRAPH

Use the graph shown in Worked example 9.3.1 to answer the following questions.

a What is the velocity of the cyclist between E and F?

b Describe the motion of the cyclist between D and E.

VELOCITY–TIME (v–t) GRAPHS

Analysing motion

A graph of velocity, v, against time, t, shows how the velocity of an object changes with time. This type of graph is useful for analysing the motion of an object moving in a complex manner.

Consider the example of the girl in Figure 9.3.5. Aliyah is running back and forth along an aisle in a supermarket. A study of the velocity–time graph in Figure 9.3.5 reveals that Aliyah is moving with a positive velocity, i.e. in a positive direction, for the first 6 s. Between the 6 s mark and the 7 s mark she is stationary, then she runs in the reverse direction, i.e. has negative velocity, for the final 3 s.

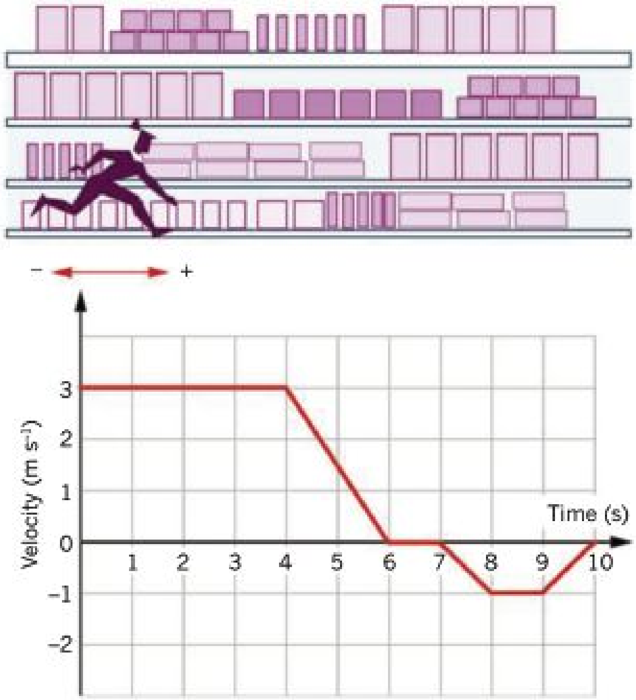


FIGURE 9.3.5 Diagram and v–t graph for a girl running along an aisle.

The graph shows Aliyah's velocity at each instant in time. She moves in a positive direction with a constant speed of 3 m s^{-1} for the first 4 s. From 4 s to 6 s, she continues moving in a positive direction but slows down. At 6 s, she comes to a stop for 1 second. During the final 3 s, she accelerates in the negative direction for 1 s then travels at a constant velocity of –1 m s^{-1} for 1 s. She then slows down and comes to a stop at 10 s. Remember that whenever the graph is below the time axis, velocity is negative, which indicates travel in the reverse direction. So she is travelling in the reverse direction for the last 3 s of her journey.

Finding displacement

A velocity–time graph can also be used to find the displacement of the object under consideration.

Displacement, s, is given by the area under a velocity–time graph, i.e. the area between the graph and the time axis. It is important to note that an area below the time axis indicates a negative displacement, i.e. motion in a negative direction.

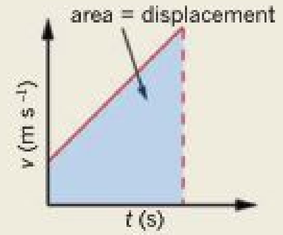


FIGURE 9.3.6 The area under a v–t graph gives displacement.

It is easier to see why the displacement is given by the area under the v–t graph when velocity is constant. For example, the graph in Figure 9.3.7 shows that in the first 6 s of motion, Aliyah moves with a constant velocity of +3 m s^{-1} for 4 s. Note that the area under the graph for this period of time is a rectangle. Her displacement, s, during this time can be determined by rearranging the formula for velocity:

$$
\begin{aligned}
v &= \frac{s}{\Delta t} \\
\therefore s &= v \times \Delta t \\
&= \text{height} \times \text{base} \\
&= \text{area under } v\text{–}t \text{ graph}
\end{aligned}
$$

Aliyah then slows from 3 m s^{-1} to zero in the next 2 s. In order to understand why the displacement for this period of time is given by the triangular area under the graph requires more complicated mathematics known as calculus, which is beyond the scope of this book.

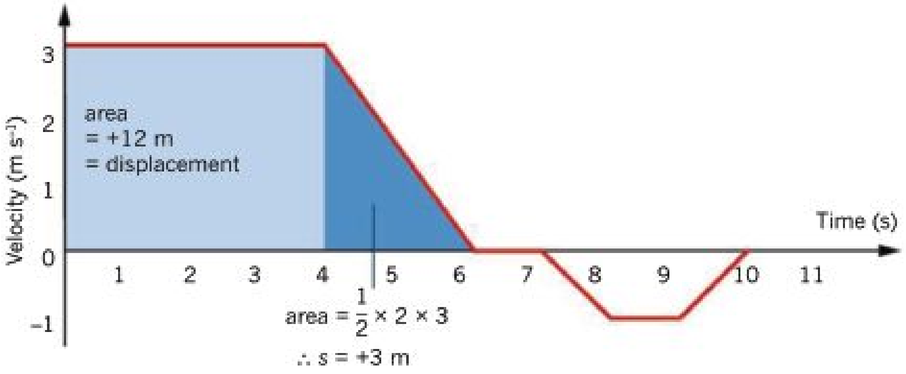


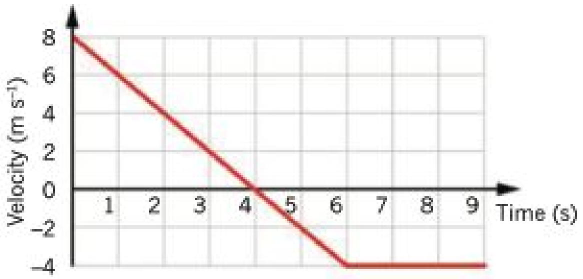
FIGURE 9.3.7 Area values as shown in a v–t graph.

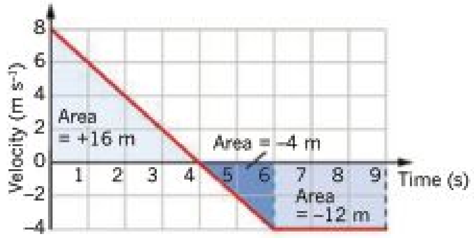
From Figure 9.3.7, the area under the graph for the first 4 s gives Aliyah's displacement during this time, i.e. +12 m. The displacement from 4 s to 6 s is represented by the area of the darker blue triangle and is equal to +3 m. The total displacement during the first 6 s is +12 m + 3 m = +15 m.

Worked example 9.3.2

ANALYSING A VELOCITY–TIME GRAPH

The motion of a radio-controlled car initially travelling east in a straight line across a driveway is represented by the graph below.



<table>
<tr><td colspan="2">a What is the displacement of the car during the first 4 seconds?</td></tr>
<tr><td>Thinking</td><td>Working</td></tr>
<tr><td>Displacement is the area under the graph. So, calculate the area under the graph for the time period for which you want to find the displacement.
Use displacement = $b \times h$ for squares and rectangles.
Use displacement = $\frac{1}{2}(b \times h)$ for triangles.</td><td>

The area from 0 to 4 s is a triangle so:
area = $\frac{1}{2}(b \times h)$
$= \frac{1}{2} \times 4 \times 8$
= +16 m</td></tr>
<tr><td>Displacement is a vector quantity, so a direction is needed.</td><td>displacement = 16 m east</td></tr>
</table>

b What is the average velocity of the car for the first 4 seconds?	
Thinking	**Working**
Identify the equation and variables, and apply the sign convention.	$v = \frac{s}{\Delta t}$ $s = +16$ m $\Delta t = 4$ s
Substitute values into the equation: $v = \frac{s}{\Delta t}$	$v = \frac{s}{\Delta t}$ $= \frac{+16}{4}$ $= +4$ m s^{-1}
Velocity is a vector quantity, so a direction is needed.	$v_{av} = 4$ m s^{-1} east

Worked example: Try yourself 9.3.2

ANALYSING A VELOCITY–TIME GRAPH

Use the graph shown in Worked example 9.3.2 to answer the following questions.

a What is the displacement of the car from 4 to 6 seconds?

b What is the average velocity of the car from 4 to 6 seconds?

The gradient of a velocity–time graph gives the average acceleration of the object over the time interval.

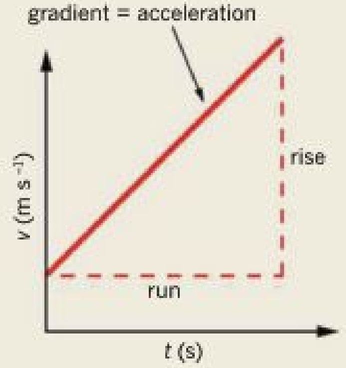


FIGURE 9.3.8 Gradient as displayed in a v–t graph.

ACCELERATION FROM A (v–t) GRAPH

The acceleration of an object can also be found from a velocity–time graph.

Consider the motion of Aliyah in the 2 s interval between 4 s and 6 s on the graph in Figure 9.3.9. She is moving in a positive direction but slowing down from 3 m s^{-1} to rest.

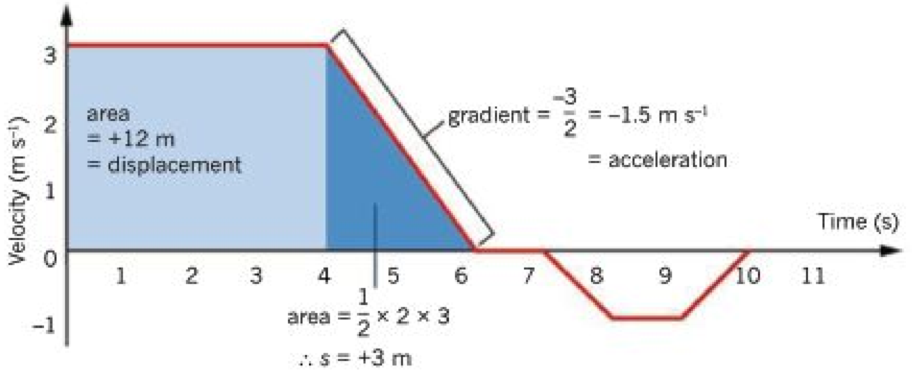


FIGURE 9.3.9 Acceleration as displayed in a v–t graph.

Her acceleration is:

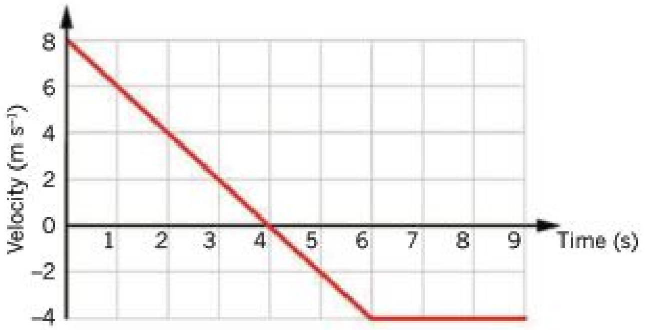
$a = \frac{\Delta v}{\Delta t} = \frac{v-u}{\Delta t} = \frac{0-3}{2} = -1.5$ m s^{-2}

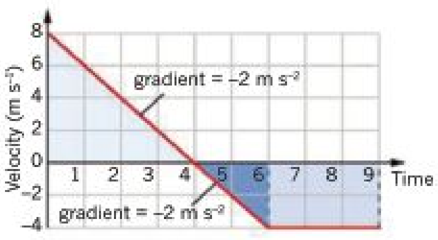
Since acceleration is the velocity change divided by time taken, it is given by the gradient of the v–t graph. As can be seen from Figure 9.3.9, the gradient of the line between 4 s and 6 s is –1.5 m s^{-2}.

Worked example 9.3.3

FINDING ACCELERATION USING A v–t GRAPH

Consider the motion of the same radio-controlled car initially travelling east in a straight line across a driveway as shown by the graph below.



What is the acceleration of the car during the first 4 s?	
Thinking	**Working**
Acceleration is the gradient of a v–t graph. Calculate the gradient using: gradient = $\frac{\text{rise}}{\text{run}}$	 Gradient from 0 – 4 = $\frac{\text{rise}}{\text{run}}$ = $\frac{-8}{4}$ = –2 m s^{-2}

Acceleration is a vector quantity, so a direction is needed. Note: In this case, the car is moving in the easterly direction and slowing down.	Acceleration = -2 m s^{-2} east (or 2 m s^{-2} west)

Worked example: Try yourself 9.3.3

FINDING ACCELERATION USING A *v*–*t* GRAPH

Use the graph shown in Worked example 9.3.3 to answer the following question.

What is the acceleration of the car during the period from 4 to 6 seconds?

DISTANCE TRAVELLED

A velocity–time graph can also be used to calculate the distance travelled by a moving object. The process of determining distance requires you to calculate the area under the v–t graph, similar to when calculating displacement. However, since distance travelled by an object always increases as the object moves, regardless of direction, you must add up all the areas between the graph and the time axis, regardless of whether the area is above or below the axis.

For example, Figure 9.3.10 shows the velocity–time graph of the radio-controlled car from Worked example 9.3.3. The area above the time-axis, which corresponds to motion in the positive direction, is +16 m, while the area below the axis, which corresponds to negative motion, consists of −4 m and −12 m. To calculate the total displacement, you would add up each displacement:

$$
\begin{aligned}
\text{total displacement} &= 16 + (-4) + (-12)\\
&= 16 - 16\\
&= 0\ \text{m}
\end{aligned}
$$

To calculate the total distance, you would add up the magnitude of the areas, by ignoring whether they are positive or negative:

$$
\begin{aligned}
\text{total distance} &= 16 + 4 + 12\\
&= 32\ \text{m}
\end{aligned}
$$

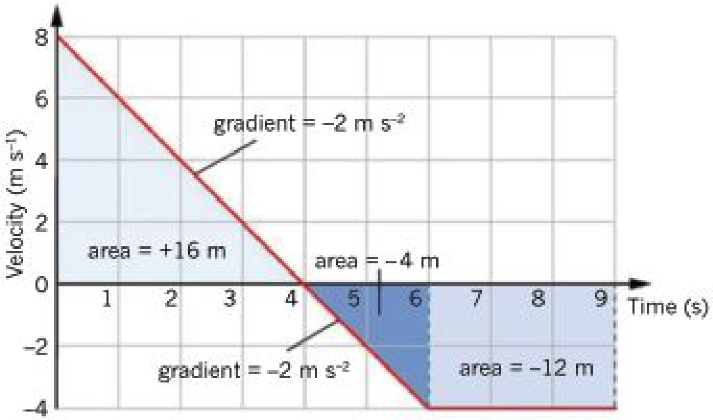


FIGURE 9.3.10 Both distance and displacement can be calculated by using the areas under the velocity–time graph.

Non-uniform acceleration

For motion with uniform (constant) acceleration, the velocity–time graph will be a straight line. For non-uniform acceleration the velocity–time graph will be curved. If the velocity–time graph is curved, the instantaneous acceleration will be the gradient of the tangent to the line at the point of interest; the average acceleration will be the gradient of the chord between two points. The displacement can still be calculated by finding the area under the graph, however you will need to make some estimations.

ACCELERATION–TIME (a–t) GRAPHS

An acceleration–time graph simply indicates the acceleration of the object as a function of time. The area under an acceleration–time graph is found by multiplying an acceleration, a, and a period of time, Δt, value. The area gives a change in velocity, Δv, value:

area $= a \times \Delta t = \Delta v$

In order to establish the actual velocity of the object, its initial velocity must be known. Figure 9.3.11 shows both Aliyah's velocity versus time (v–t) and acceleration versus time (a–t) graphs.

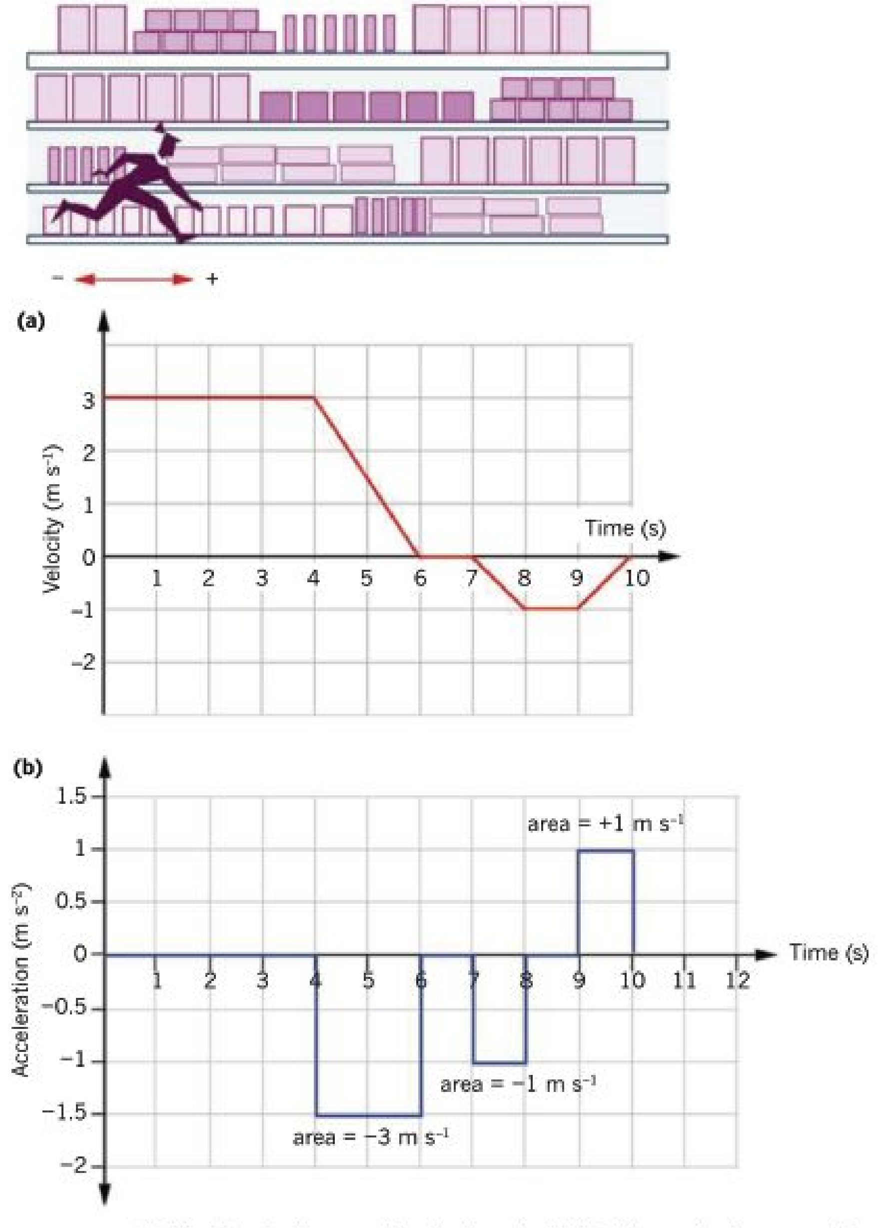


FIGURE 9.3.11 (a) Aliyah's velocity versus time (v–t) graph. (b) Aliyah's acceleration versus time (a–t) graph.

From 4 to 6 s, the area shows a Δv of -3 m s^{-1}. This indicates that she has slowed down by 3 m s^{-1} during this time. Her v–t graph confirms this fact. Her initial speed is 3 m s^{-1}, so she must be stationary ($v = 0$) after 6 s. This calculation could not be made without knowing the girl's initial velocity.

9.3 Review

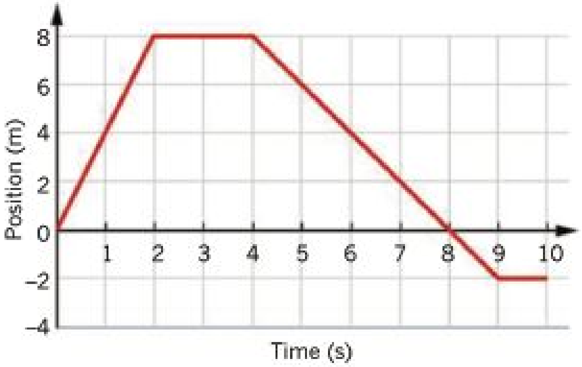
SUMMARY

- A position–time graph can be used to determine the location of an object at any given time. Additional information can also be derived from the graph:
  - Displacement is given by the change in position of an object.
  - The velocity of an object is given by the gradient of the position–time graph.
  - If the position–time graph is curved, the gradient of the tangent at a point gives the instantaneous velocity.
- The gradient of a velocity–time graph is the acceleration of the object.
- The area under a velocity–time graph is the displacement of the object.
- The area under an acceleration–time graph is the change in velocity of the object.

KEY QUESTIONS

1 Which of the following does the gradient of a position–time graph represent?
A displacement
B acceleration
C time
D velocity

The following information relates to questions 2–6. The graph represents the straight-line motion of a radio-controlled toy car.



2 Describe the motion of the car in terms of its position.

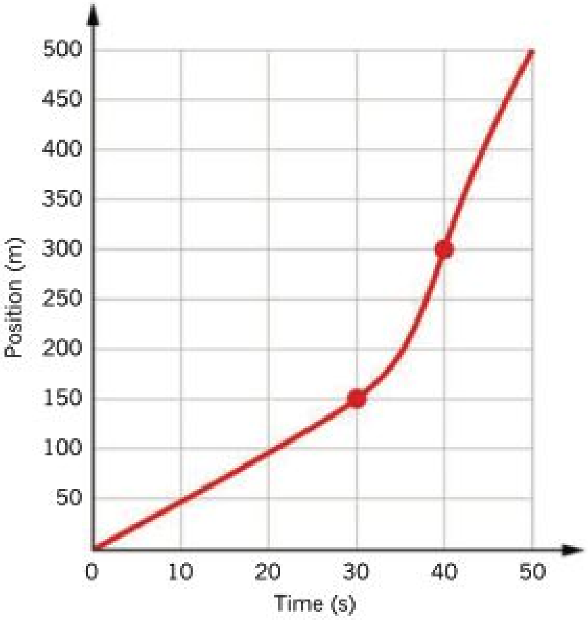
3 What was the position of the toy car after:
a 2 s?
b 4 s?
c 6 s?
d 10 s?

4 When did the car return to its starting point?

5 What was the velocity of the toy car:
a during the first 2 s?
b at 3 s?
c from 4 s to 8 s?
d at 8 s?
e from 8 s to 9 s?

6 During its 10 s motion, what was the car's:
a distance travelled?
b displacement?

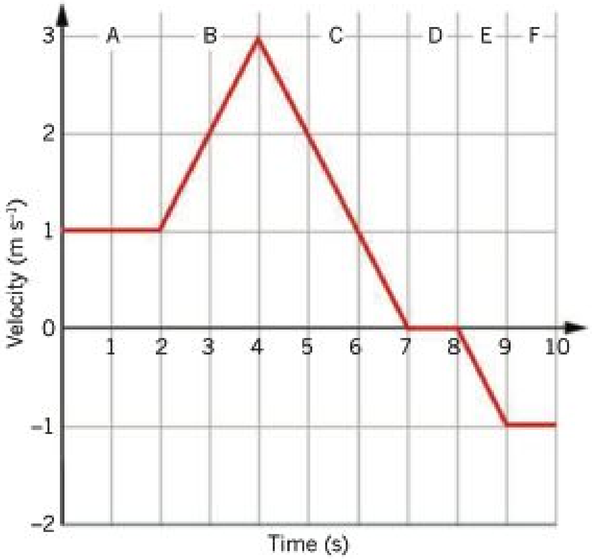
7 This position–time graph for a cyclist travelling north along a straight road is shown. Calculate the following information about the cyclist's motion.



a What was the average speed of the cyclist during the first 30 s?
b What was the average velocity of the cyclist during the final 10 s?
c What was the average velocity of the cyclist for the whole trip?

9.3 Review *continued*

8 The graph in the figure below shows the motion of a dog running along a footpath.



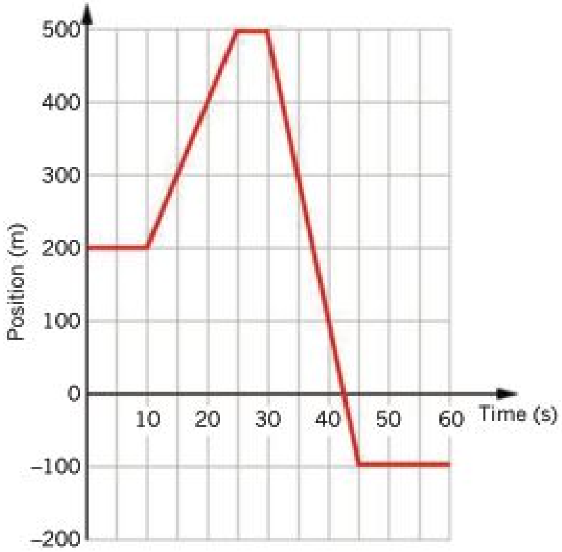
a What is the magnitude of the acceleration of the dog at $t = 1$ s?

b What is the magnitude of the acceleration of the dog at $t = 5$ s?

c What is the magnitude of the displacement of the dog for the first 7 s?

d What is the magnitude of the average velocity of the dog over the first 7 s?

9 The graph shows the position of a motorbike along a straight stretch of road as a function of time. The motorcyclist starts 200 m north of an intersection.

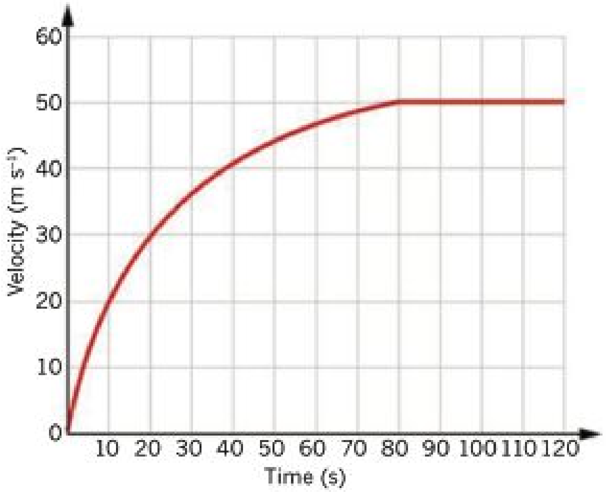


Calculate the instantaneous velocity of the motorcyclist at each of the following times:

a 15 s

b 35 s.

10 The straight-line motion of a high-speed intercity train is shown in the graph below.



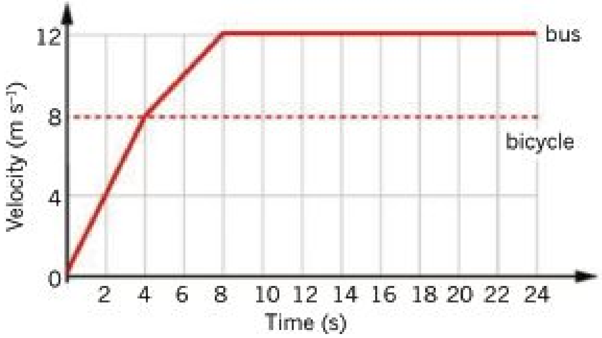
a How long does it take the train to reach its cruising speed?

b What is the acceleration of the train 10 s after starting?

c What is the acceleration of the train 40 s after starting?

d By counting squares, or by another suitable method, approximate the displacement (in km) of the train after 120 s.

11 The velocity–time graphs for a bus and a bicycle travelling along the same straight stretch of road are shown below. The bus is initially at rest and starts moving as the bicycle passes it.



a What is the magnitude of the initial acceleration of the bus?

b At what time does the bus overtake the bicycle?

c How far has the bicycle travelled before the bus catches it?

d What is the magnitude of the average velocity of the bus during the first 8 s?

12 a Draw an acceleration–time graph for the bus discussed in question 11.

b Use your acceleration–time graph to determine the change in velocity of the bus over the first 8 s.

9.4 Equations for uniform acceleration

A graph is an excellent way of representing motion because it provides a great deal of information that is easy to interpret. However, a graph is time-consuming to draw and sometimes values have to be estimated rather than precisely calculated.

In the previous section, graphs of motion were used to determine quantities such as displacement, velocity and acceleration. This section examines a more powerful and precise method of solving problems involving *constant* or *uniform acceleration*. This method involves the use of a series of equations that can be derived from the basic definitions developed earlier.

DERIVING THE EQUATIONS

Consider an object moving in a straight line with an initial velocity, u, and a uniform acceleration, a, for a time interval, Δt. As u, v and a are vectors, and the motion is limited to one dimension, the sign and direction convention of right as positive and left as negative can be used. After a period of time, Δt, the object has changed its velocity from an initial velocity of u and is now travelling with a final velocity of v. Its acceleration will be given by:

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

If the initial time is 0 s, and the final time is t s, then $\Delta t = t$. The above equation can then be rearranged as:

$v = u + at$ (i)

The average velocity of the object is:

average velocity $v_{av} = \frac{\text{displacement}}{\text{time taken}} = \frac{s}{\Delta t}$

When acceleration is uniform, average velocity, v_{av}, can also be found as the average of the initial and final velocities:

$v_{av} = \frac{1}{2}(u + v)$

This relationship is shown graphically in Figure 9.4.1.

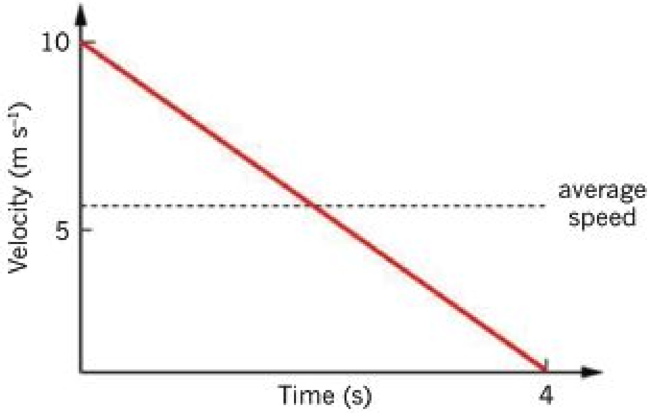


FIGURE 9.4.1 Uniform acceleration as displayed by a v–t graph.

So:

$\frac{s}{t} = \frac{1}{2}(u + v)$

This gives:

$s = \frac{1}{2}(u + v)t$ (ii)

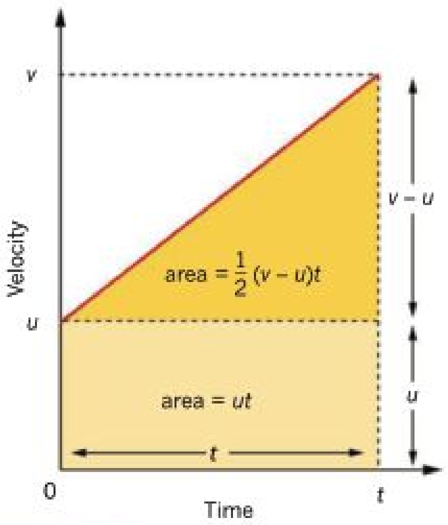


FIGURE 9.4.2 The area under a v–t graph broken up into a rectangle and a triangle.

A graph describing constant acceleration motion is shown in Figure 9.4.2. For constant acceleration, the velocity is increasing by the same amount in each time interval, so the gradient of the v–t graph is constant.

The displacement, s, of the body is given by the area under the velocity–time graph. The area under the velocity–time graph, as shown in Figure 9.4.2, is given by the combined area of the rectangle and the triangle:

Area $= s = ut + \frac{1}{2} \times (v - u) \times t$

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

then: $v - u = at$, and this can be substituted for $v - u$:

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

$$s = ut + \frac{1}{2}at^2 \quad \text{(iii)}$$

Making u the subject of equation (i) gives:

$u = v - at$

You might like to derive another equation yourself by substituting this into equation (ii). You will get:

$$s = vt - \frac{1}{2}at^2 \quad \text{(iv)}$$

Rewriting equation (i) with t as the subject gives:

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

Now, if this is substituted into equation (ii):

$$\begin{aligned} s &= \frac{1}{2}(u + v)t \\ &= \frac{u + v}{2} \times \frac{v - u}{a} \\ &= \frac{v^2 - u^2}{2a} \end{aligned}$$

Finally, transposing this gives:

$$v^2 = u^2 + 2as \quad \text{(v)}$$

Equations (i)–(v) are commonly used to solve problems in which acceleration is constant. They are summarised below.

$v = u + at$

$s = \frac{1}{2}(u + v)t$

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

$s = vt - \frac{1}{2}at^2$

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

where s is the displacement (in m)

u is the initial velocity (in m s^{-1})

v is the final velocity (in m s^{-1})

a is the acceleration (in m s^{-2})

t is the time taken (in s).

SOLVING PROBLEMS USING EQUATIONS

When solving problems using these equations, it is important to think about the problem and try to visualise what is happening. Follow the steps below.

Step 1 Draw a simple diagram of the situation.

Step 2 Write down the information that has been given in the question. You might like to use the word 'suvat' as a memory trick to help you remember to list the variables in the order *s*, *u*, *v*, *a* and *t*. Use a sign convention to assign positive and negative values to indicate directions. Convert all units to SI form.

Step 3 Select the equation that matches your data. It should include three values that you know, and the one value that you want to solve.

Step 4 Use the appropriate number of significant figures in your answer.

Step 5 Include units with the answer and specify a direction if the quantity is a vector.

Worked example 9.4.1

USING THE EQUATIONS OF MOTION

A snowboarder in a race is travelling 10 m s^{-1} north as he crosses the finishing line. He then decelerates uniformly, coming to a stop over a distance of 20 m.

a What is his acceleration as he comes to a stop?	
Thinking	**Working**
Write down the known quantities as well as the quantity you are finding. Apply the sign convention that north is positive and south is negative.	Take all the information that you can from the question: • constant acceleration, so use equations for uniform acceleration • 'coming to a stop' means that the final velocity is zero. $s = +20$ m $u = +10$ m s^{-1} $v = 0$ m s^{-1} $a = ?$
Identify the correct equation to use.	$v^2 = u^2 + 2as$
Substitute known values into the equation and solve for *a*. Include units with the answer.	$v^2 = u^2 + 2as$ $0^2 = 10^2 + 2 \times a \times 20$ $0 = 100 + 40a$ $-100 = 40a$ $a = \frac{-100}{40}$ $= -2.5$ m s^{-2}
Use the sign convention to state the answer with its direction.	$a = 2.5$ m s^{-2} south

b How long does he take to come to a stop?

Thinking	Working
Write down the known quantities as well as the quantity you are finding. Apply the sign convention that north is positive and south is negative.	Take all the information that you can from the question: • constant acceleration, so use equations for uniform acceleration • 'coming to a stop' means that the final velocity is zero. $s = +20$ m $u = +10$ m s^{-1} $v = 0$ m s^{-1} $a = -2.5$ m s^{-2} $t = ?$
Identify the correct equation to use. Since you now know four values, any equation involving *t* will work.	$v = u + at$
Substitute known values into the equation and solve for *t*. Include units with the answer.	$v = u + at$ $0 = 10 + (-2.5) \times t$ $-10 = -2.5t$ $t = \frac{-10}{-2.5}$ $= 4.0$ s

c What is the average velocity of the snowboarder as he comes to a stop?

Thinking	Working
Write down the known quantities as well as the quantity that you are finding. Apply the sign convention that north is positive and south is negative.	Take all the information that you can from the question: • constant acceleration, so we only need to find the average of the final and initial speeds. $u = +10$ m s^{-1} $v = 0$ m s^{-1} $v_{av} = ?$
Identify the correct equation to use.	$v_{av} = \frac{1}{2}(u + v)$
Substitute known values into the equation and solve for v_{av}. Include units with the answer.	$v_{av} = \frac{1}{2}(u + v)$ $= \frac{1}{2}(0 + 10)$ $= 5.0$ m s^{-1}
Use the sign convention to state the answer with its direction.	$v_{av} = 5.0$ m s^{-1} north

Worked example: Try yourself 9.4.1

USING THE EQUATIONS OF MOTION

A snowboarder in a race is travelling 15 m s^{-1} east as she crosses the finishing line. She then decelerates uniformly until coming to a stop over a distance of 30 m.

a What is her acceleration as she comes to a stop?

b How long does she take to come to a stop?

c What is the average velocity of the snowboarder as she comes to a stop?

9.4 Review

SUMMARY

- The following equations can be used for situations where there is a constant acceleration, where:
  - s = displacement (m)
  - u = initial velocity (m s^{-1})
  - v = final velocity (m s^{-1})
  - a = acceleration (m s^{-2})
  - t = time (s).
- $v = u + at$
- $s = \frac{1}{2}(u + v)t$
- $s = ut + \frac{1}{2}at^2$
- $s = vt - \frac{1}{2}at^2$
- $v^2 = u^2 + 2as$
- $v_{av} = \frac{s}{t} = \frac{u + v}{2}$
- A sign and direction convention for motion in one dimension needs to be used with these equations.

KEY QUESTIONS

1 A cyclist has a uniform acceleration as he rolls down a hill. His initial speed is 5 m s^{-1}, he travels a distance of 30 m and his final speed is 18 m s^{-1}. Which equation should be used to determine his acceleration?

A $v = u + at$

B $s = \frac{1}{2}(u + v)t$

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

D $s = vt - \frac{1}{2}at^2$

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

2 A new-model Subaru travels with a uniform acceleration on a racetrack. It starts from rest and covers 400 m in 16 s.

a What is the magnitude of its average acceleration during this time?

b What is the final speed of the car in m s^{-1}?

c What is the car's final speed in km h^{-1}?

3 A Prius hybrid car starts from rest and accelerates uniformly in a positive direction for 8.0 s. It reaches a final speed of 16 m s^{-1}.

a What is the magnitude of the acceleration of the Prius?

b What is the magnitude of the average velocity of the Prius?

c What is the distance travelled by the Prius?

4 During its launch phase, a space rocket accelerates uniformly from rest to 160 m s^{-1} upwards in 4.0 s, then travels with a constant speed of 160 m s^{-1} for the next 5.0 s.

a What is the initial acceleration of the rocket?

b How far (in km) does the rocket travel in this 9.0 s period?

c What is the final speed of the rocket in km h^{-1}?

d What is the average speed of the rocket during the first 4.0 s?

e What is the average speed of the rocket during the 9.0 s motion?

5 While overtaking another cyclist, Ben increases his speed uniformly from 4.2 m s^{-1} to 6.7 m s^{-1} east over a time interval of 0.50 s.

a What is the magnitude of Ben's average acceleration during this time?

b How far does Ben travel while overtaking?

c What is Ben's average speed during this time?

6 A stone is dropped vertically into a lake. Which one of the following statements best describes the motion of the stone at the instant it enters the water?

A Its velocity and acceleration are both downwards.

B It has an upwards velocity and a net downwards acceleration.

C Its velocity and acceleration are both upwards.

D It has a downwards velocity and a net upwards acceleration.

7 A diver plunges headfirst into a diving pool while travelling at 28 m s^{-1} downwards. The diver enters the water and stops after a distance of 4.0 m. Consider the diver to be a single point located at her centre of mass and assume her acceleration through the water to be uniform.

a What is the magnitude of the average acceleration of the diver as she travels through the water?

b How long does the diver take to come to a stop?

c What is the speed of the diver after she has dived through 2.0 m of water?

8 A car is travelling along a straight road at 75 km h^{-1} east. In an attempt to avoid an accident, the motorist has to brake suddenly and stop the car.

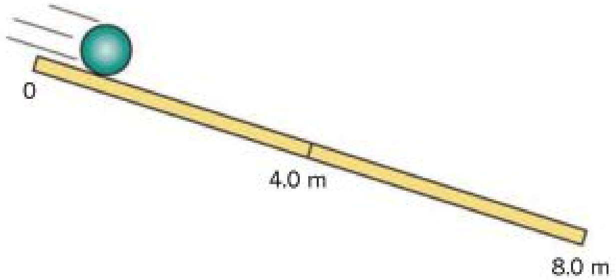
a What is the car's initial speed in m s^{-1}?

b If the reaction time of the motorist is 0.25 s, what distance does the car travel while the driver is reacting to apply the brakes?

9.4 Review *continued*

- **c** Once the brakes are applied, the car has an acceleration of $-6.0\ \text{m s}^{-2}$. How far does the car travel while pulling up?
- **d** What total distance does the car travel from the time the driver first notices the danger to when the car comes to a stop?

9 A billiard ball rolls from rest down a smooth ramp that is 8.0 m long. The acceleration of the ball is constant at $2.0\ \text{m s}^{-2}$.



- **a** What is the speed of the ball when it is halfway down the ramp?
- **b** What is the final speed of the ball?
- **c** How long does the ball take to roll the first 4.0 m?
- **d** How long does the ball take to travel the final 4.0 m?

10 A cyclist, Anna, is travelling at a constant speed of $12\ \text{m s}^{-1}$ when she passes a stationary bus. The bus starts moving just as Anna passes, and it accelerates uniformly at $1.5\ \text{m s}^{-2}$.

- **a** When does the bus reach the same speed as Anna?
- **b** How long does the bus take to catch Anna?
- **c** What distance has Anna travelled before the bus catches up?

9.5 Vertical motion

Until 500 years ago, it was widely believed that the heavier an object was, the faster it would fall. This was the theory of Aristotle, and it lasted for 2000 years until the end of the Middle Ages. In the seventeenth century, the Italian scientist Galileo conducted experiments that showed that the mass of the object did not affect the rate at which it fell, as long as air resistance was not a factor.

It is now known that falling objects speed up because of gravity. Many people still think that heavier objects fall faster than light objects. This is not the case, but the confusion arises because of the effects of air resistance. This section examines the motion of falling objects.

ANALYSING VERTICAL MOTION

Some falling objects are affected by **air resistance** to a large extent, for example, feathers and balloons. This is why these objects do not speed up much as they fall. However, if air resistance can be ignored, all **free-falling** bodies near the Earth's surface will move with an equal downwards acceleration. The stroboscopic image in Figure 9.5.1 clearly shows an apple accelerating as it falls, since the distance travelled by the apple between each photograph increases. In a vacuum, this rate of acceleration would be the same for a feather, a bowling ball, or any other object. The mass of the object does not matter.

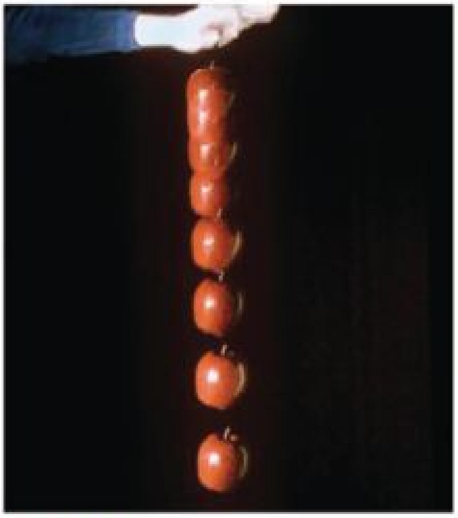


FIGURE 9.5.1 A stroboscopic image of a free-falling apple. The time elapsed between each image of the apple is the same but the distance it travels increases, which shows the apple is accelerating. Without air resistance, this rate of acceleration is the same for all objects.

At the Earth's surface, the acceleration due to gravity, g, is 9.8 m s^{-2} down and does not depend on whether the body has been thrown upwards or is falling down.

As an example, a coin that is dropped from rest will be moving at 9.8 m s^{-1} after 1 s, 19.6 m s^{-1} after 2 s, and so on. Each second its speed increases by 9.8 m s^{-1}. The motion of a falling coin is illustrated in Figure 9.5.2.

However, if the coin was launched straight up at 19.6 m s^{-1}, then after 1 s its speed would be 9.8 m s^{-1}, and after 2 s it would be stationary. In other words, each second it would slow down by 9.8 m s^{-1}. The motion of a coin thrown vertically upwards is shown in Figure 9.5.3.

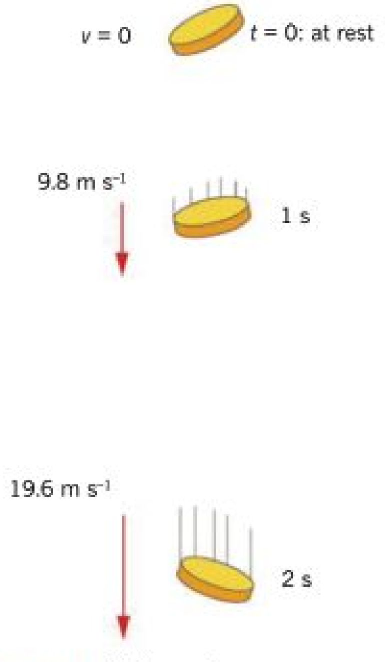


FIGURE 9.5.2 A falling coin.

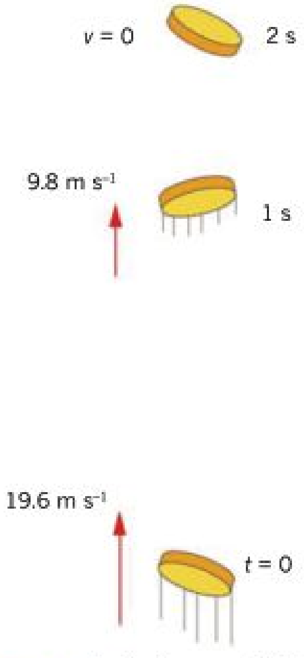


FIGURE 9.5.3 A coin thrown vertically upwards.

So, regardless of whether the coin is falling or is tossed, its speed changes at the same rate. The speed of the falling coin *increases* by 9.8 m s^{-1} each second and the speed of the rising coin *decreases* by 9.8 m s^{-1} each second. That means that the acceleration of the coin due to gravity is 9.8 m s^{-2} downwards in both cases.

PHYSICSFILE

Galileo's experiment on the Moon

In 1971, David Scott went to great lengths to show that Galileo's prediction was correct. As an astronaut on the Apollo 15 Moon mission, he took a hammer and a feather on the voyage. He stepped onto the lunar surface, held the feather and hammer at the same height and dropped them together. As Galileo had predicted 400 years earlier, in the absence of any air resistance the two objects fell side by side as they accelerated towards the Moon's surface.

FIGURE 9.5.4 Astronaut David Scott holding a feather and a hammer on the Moon.

PHYSICS IN ACTION

Theories of motion: Aristotle and Galileo

Aristotle was a Greek philosopher who lived in the fourth century BCE. He was such an influential individual that his ideas on motion were generally accepted for nearly 2000 years. Aristotle did not do experiments as we know them today, but simply *thought* about different bodies in order to arrive at a plausible explanation for their motion.

Aristotle spent a lot of time classifying animals, and adopted a similar approach in his study of motion. His theory gave inanimate objects, such as rocks and rain, similar characteristics to living things. Aristotle organised objects into four terrestrial groups or elements: earth, water, air and fire (see Figure 9.5.5). He said that any object was a mixture of these elements in a certain proportion.

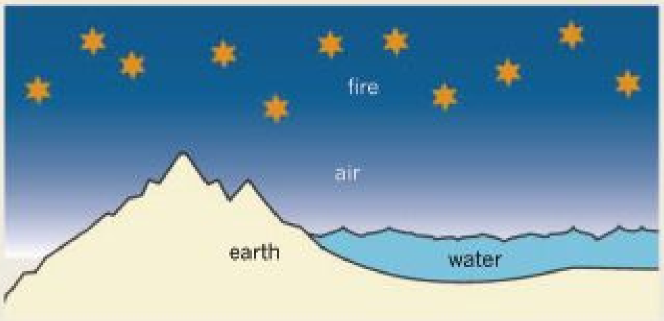


FIGURE 9.5.5 Aristotle's four elements of the universe; earth, water, air and fire.

According to Aristotle, a body would move because of a tendency that could come from inside or outside the body. An internal tendency would cause 'natural' motion and result in a body returning to its proper place. For example, if a rock, which is an earth substance, is held in the air and released, its natural tendency would be to return to Earth. This explains why it falls down. Similarly, fire was thought to head upwards in an attempt to return to its proper place in the universe.

An external push that acts when something is thrown or hit was the cause of 'violent' motion in the Aristotelian model. An external push acted to take a body away from its proper place. For example, when an apple is thrown into the air, a violent motion carries the apple away from the Earth, but then the natural tendency of the apple takes over and it returns to its home.

Aristotle's theory worked quite well and could be used to explain many observed types of motion. However, there were also many examples that it could not successfully explain, such as why some solids floated instead of sinking.

Aristotle explained the behaviour of a falling body by saying that its speed depended on how much earth element it contained. This suggested that a 2 kg cat would fall twice as fast and in half the time as a 1 kg cat dropped from the same height. Many centuries later, Galileo Galilei (pictured in Figure 9.5.6) noticed that, at the start of a hailstorm, small hailstones arrived at the same time as large hailstones. This caused Galileo to doubt Aristotle's theory and so he set about finding an explanation for the motion of freely falling bodies.

FIGURE 9.5.6 Galileo Galilei.

A famous story in science is that of Galileo dropping different weights from the Leaning Tower of Pisa in Italy. This story may or may not be true, but Galileo did perform a very detailed analysis of falling bodies. Galileo used inclined planes because freely falling bodies moved too fast to analyse. He completed extensive and thorough experiments that showed conclusively that Aristotle was incorrect.

By using a water clock to time balls as they rolled down different inclines, he was able to show that the balls were accelerating and that the distance they travelled was proportional to the square of the time, i.e. $d \propto t^2$.

Galileo found that this also held true when he inclined the plane at larger and larger angles, allowing him to conclude that freely falling bodies actually fall with a uniform acceleration.

PHYSICSFILE

Strength of gravity

The acceleration due to gravity, g, on Earth varies slightly from 9.8 m s^{-2} according to the location. The reasons for this will be studied in Physics Unit 3. On the Moon, the strength of gravity, g, is much weaker than on Earth and falling objects accelerate at 1.6 m s^{-2}. Other planets and bodies in the solar system have different values of g depending on their mass and size. The value of g at various locations is provided in Table 9.5.1.

Location	Acceleration due to gravity (m s^{-2})
Melbourne	9.800
South Pole	9.832
Equator	9.780
Moon	1.600
Mars	3.600
Jupiter	24.600
Pluto	0.670

TABLE 9.5.1 Acceleration due to gravity at different locations on Earth, and on other bodies in the solar system.

In Physics Unit 2 you can assume that acceleration due to gravity is always 9.8 m s^{-2}.

Since the acceleration of a free-falling body is constant, it is appropriate to use the equations that were studied in the previous section, 'Equations for uniform acceleration'. It is necessary to specify whether up or down is positive when doing these problems. You can simply follow the mathematical convention of regarding up as positive, which would mean the acceleration due to gravity would always be –9.8 m s^{-2}.

Worked example 9.5.1

VERTICAL MOTION

A construction worker accidentally knocks a brick from a building so that it falls vertically a distance of 50 m to the ground. Use $g = -9.8$ m s^{-2} and ignore air resistance when answering these questions.

a How long does the brick take to fall halfway, to 25 m?	
Thinking	**Working**
Write down the values of the quantities that are known and what you are finding. Apply the sign convention that up is positive and down is negative.	The brick starts at rest so $u = 0$. $s = -25$ m $u = 0$ m s^{-1} $a = -9.8$ m s^{-2} $t = ?$
Select the equation for uniform acceleration that best fits the data you have.	$s = ut + \frac{1}{2}at^2$
Substitute known values into the equation and solve for t. Think about whether the value seems reasonable.	$-25 = 0 \times t + \frac{1}{2} \times -9.8 \times t^2$ $= -4.9t^2$ $t = \sqrt{\frac{-25}{-4.9}}$ $= 2.3$ s

b How long does the brick take to fall all the way to the ground?

Thinking	Working
Write down the values of the quantities that are known and what you are finding. Apply the sign convention that up is positive and down is negative.	$s = –50$ m $u = 0$ m s^{-1} $a = –9.8$ m s^{-1} $t = ?$
Identify the correct equation of uniform acceleration to use.	$s = ut + \frac{1}{2}at^2$
Substitute known values into the equation and solve for *t*. Think about whether the value seems reasonable. Notice that the brick takes 2.3 s to travel the first 25 m and only 0.9 s to travel the final 25 m. This is because it is accelerating.	$-50 = 0 \times t + \frac{1}{2} \times -9.8 \times t^2$ $-50 = -4.9t^2$ $t = \sqrt{\frac{-50}{-4.9}}$ $= 3.2$ s

c What is the speed of the brick as it hits the ground?

Thinking	Working
Write down the values of the quantities that are known and what you are finding. Apply the sign convention that up is positive and down is negative	$s = –50$ m $u = 0$ m s^{-1} $v = ?$ $a = –9.8$ m s^{-1} $t = 3.2$ s
Identify the correct equation to use. Since you now know four values, any equation involving *v* will work.	$v = u + at$
Substitute known values into the equation and solve for *v*. Think about whether the value seems reasonable.	$v = 0 + (-9.8) \times 3.2$ $= -31$ m s^{-1}
Use the sign and direction convention to describe the direction of the final velocity.	$v = -31$ m s^{-1} or 31 m s^{-1} downwards

Worked example: Try yourself 9.5.1

VERTICAL MOTION

A construction worker accidentally knocks a hammer from a building so that it falls vertically a distance of 60 m to the ground. Use $g = -9.8$ m s^{-2} and ignore air resistance when answering these questions.

a How long does the hammer take to fall halfway, to 30 m?

b How long does it take the hammer to fall all the way to the ground?

c What is the speed of the hammer as it hits the ground?

When an object is thrown vertically up into the air, it will eventually reach a point where it stops momentarily before returning back down. So the velocity of the object decreases as the object rises, becomes zero at the maximum height, and then increases again as the object falls. Throughout this motion, however, the object is still in the same gravitational field, so *g* remains at –9.8 m s^{-2} throughout the journey. Knowing that the velocity of an object thrown in the air is zero at the top of its flight allows you to calculate the maximum height reached.

Worked example 9.5.2

MAXIMUM HEIGHT PROBLEMS

On winning a tennis match the victorious player, Michael, smashed the ball vertically into the air at 27.5 m s^{-1}. In this example, air resistance can be ignored and the acceleration due to gravity will be taken as 9.8 m s^{-2}.

a Determine the maximum height reached by the ball.	
Thinking	**Working**
Write down the values of the quantities that are known and what you are finding. At the maximum height the velocity is zero. Apply the sign convention that up is positive and down is negative.	$u = 27.5$ m s^{-1} $v = 0$ $a = -9.8$ m s^{-2} $s = ?$
Select an appropriate formula.	$v^2 = u^2 + 2as$
Substitute known values into the equation and solve for *s*.	$0 = (27.5)^2 + 2 \times (-9.8) \times s$ $s = \frac{-756.25}{-19.6}$ $\therefore s = +38.6$ m, i.e. the ball reaches a height of 38.6 m.

b Calculate the time that the ball takes to return to its starting position.	
Thinking	**Working**
To work out the time for which the ball is in the air, it is often necessary to first calculate the time that it takes to reach its maximum height. Write down the values of the quantities that are known and what you are finding.	$u = 27.5$ m s^{-1} $v = 0$ m s^{-1} $a = -9.8$ m s^{-2} $s = 38.6$ $t = ?$
Select an appropriate formula.	$v = u + at$
Substitute known values into the equation and solve for *t*.	$0 = 27.5 + (-9.8 \times t)$ $9.8t = 27.5$ $\therefore t = 2.8$ s The ball takes 2.8 s to reach its maximum height. It will therefore take 2.8 s to fall from this height back to its starting point and so the whole trip will last for 5.6 s.

Worked example: Try yourself 9.5.2

MAXIMUM HEIGHT PROBLEMS

On winning a cricket match, a fielder throws a cricket ball vertically into the air at 15 m s^{-1}. In this example, air resistance can be ignored and the acceleration due to gravity will be taken as 9.8 m s^{-2}.

a Determine the maximum height reached by the ball.

b Calculate the time that the ball takes to return to its starting position.

9.5 Review

SUMMARY

- If air resistance can be ignored, all bodies falling freely near the Earth will move with the same constant acceleration.
- The acceleration due to gravity is represented by *g* and is equal to 9.8 m s^{-2} in the direction towards the centre of the Earth.
- The equations for uniform acceleration can be used to solve vertical motion problems. It is necessary to specify whether up or down is positive.

KEY QUESTIONS

For these questions, ignore the effects of air resistance and assume that the acceleration due to gravity is 9.8 m s^{-2} unless instructed otherwise.

1 Angus inadvertently drops an egg while baking a cake and the egg falls vertically towards the ground. Which one of the following statements correctly describes how the egg falls?

A The egg's acceleration increases.
B The egg's acceleration is constant.
C The egg's velocity is constant.
D The egg's acceleration decreases.

2 Max is an Olympic trampolinist and is practising some routines. Which one or more of the following statements correctly describes Max's motion when he is at the highest point of the bounce? Assume that his motion is vertical.

A He has zero velocity.
B His acceleration is zero.
C His acceleration is upwards.
D His acceleration is downwards.

3 A window cleaner working on the Bell tower accidently drops her mobile phone. The phone falls vertically towards the ground with an acceleration of 9.8 m s^{-2}.

a Determine the speed of the phone after 3.0 s.
b How fast is the phone moving after it has fallen 30 m?
c What is the average velocity of the phone during a fall of 30 m?

4 A girl tosses a marble straight up into the air at 5 m s^{-1} and then catches it at the same height from which it was thrown. Ignore air resistance.

a Is the acceleration of the marble on the way up the same as, less than or greater than its acceleration on the way down? Justify your answer.
b Is the launch speed of the marble the same as, less than or greater than its landing speed? Justify your answer.

5 A super ball is bounced so that it travels straight up into the air, reaching its highest point after 1.5 s.

a What is the initial speed of the ball just as it leaves the ground?
b What is the maximum height reached by the ball?

6 A book is knocked off a bench and falls vertically to the floor. If the book takes 0.40 s to fall to the floor, calculate the following descriptions of its motion.

a What is the book's speed as it lands?
b What is the height from which the book fell?
c How far did the book fall during the first 0.20 s?
d How far did the book fall during the final 0.20 s?

7 While celebrating her birthday, Bindi pops a party popper. The lid travels vertically into the air. Being a keen physics student, Bindi notices that the lid takes 4.0 s to return to its starting position.

a How long does the lid take to reach its maximum height?
b How fast was the lid travelling initially?
c What was the maximum height reached by the lid?
d What was the velocity of the lid as it returned to its starting point?

8 Two physics students conduct the following experiment from a very high bridge. Thao drops a 1.5 kg shot-put from a vertical height of 60.0 m, while at exactly the same time Benjamin throws a 100 g mass with an initial downwards velocity of 10.0 m s^{-1} from a point 10.0 m above Thao.

a How long does it take the shot-put to reach the ground?
b How long does it take the 100 g mass to reach the ground?

9 At the start of a football match, the umpire bounces the ball so that it travels vertically upwards and reaches a height of 15.0 m.

a How long does the ball take to reach this maximum height?
b One of the ruckmen is able to leap and reach to a height of 4.0 m with his hand. How long after the bounce should this ruckman try to make contact with the ball?

Chapter review

KEY TERMS

acceleration
air resistance
centre of mass
dimensional analysis
displacement
distance travelled
free fall
magnitude
position
speed
velocity

For the following questions, the acceleration due to gravity is 9.8 m s^{-2} down and air resistance is considered to be negligible, unless indicated otherwise.

1 A car travels at 95 km h^{-1} along a freeway. What is its speed in m s^{-1}?

2 A cyclist travels at 15 m s^{-1} during a sprint finish. What is this speed in km h^{-1}?

The following information relates to questions 3 and 4.
An athlete in training for a marathon runs 15 km north along a straight road before realising that she has dropped her drink bottle. She turns around and runs back 5 km to find her bottle, then resumes running in the original direction. After running for 2.0 hours, the athlete reaches 20 km from her starting position and stops.

3 Calculate the average speed of the athlete in km h^{-1}.

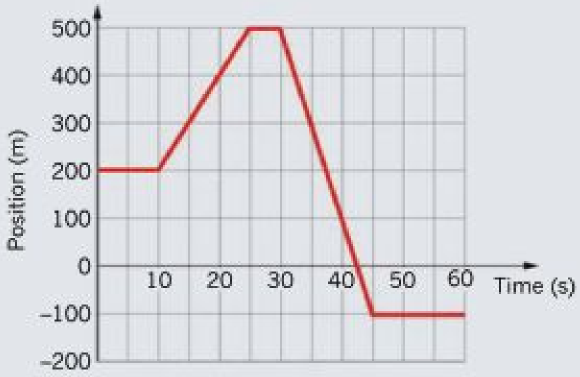
4 Calculate her average velocity in:
  a km h^{-1}
  b m s^{-1}.

5 A ping pong ball is falling vertically at 6.0 m s^{-1} as it hits the floor. It rebounds at 4.0 m s^{-1} up. What is its change in speed during the bounce?

6 A car is moving in a positive direction. It approaches a red light and slows down. Which of the following statements correctly describes its acceleration and velocity as it slows down?
  A The car has positive acceleration and negative velocity.
  B The car has negative acceleration and positive velocity.
  C Both the velocity and acceleration of the car are positive.
  D Both the velocity and acceleration of the car are negative.

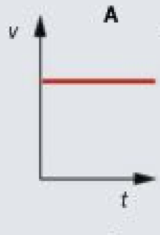
7 A skier is travelling along a horizontal ski run at a speed of 15 m s^{-1}. After falling over, the skier takes 2.5 s to come to rest. Calculate the average acceleration of the skier.

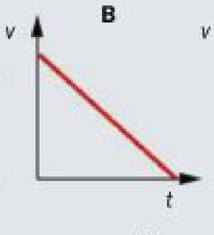
8 The graph below shows the position of a motorbike along a straight stretch of road as a function of time. The motorcyclist starts 200 m north of an intersection.

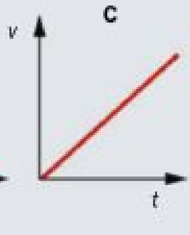


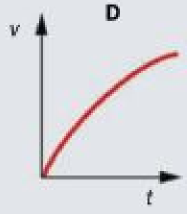
  a At what time interval is the motorcyclist travelling in a northerly direction?
  b At what time interval is the motorcyclist travelling in a southerly direction?
  c At what time intervals is the motorcyclist stationary?
  d At what time is the motorcyclist passing back through the intersection?

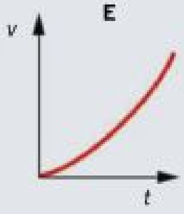
9 For each of the activities below, indicate which of the following velocity–time graphs best represents the motion involved.







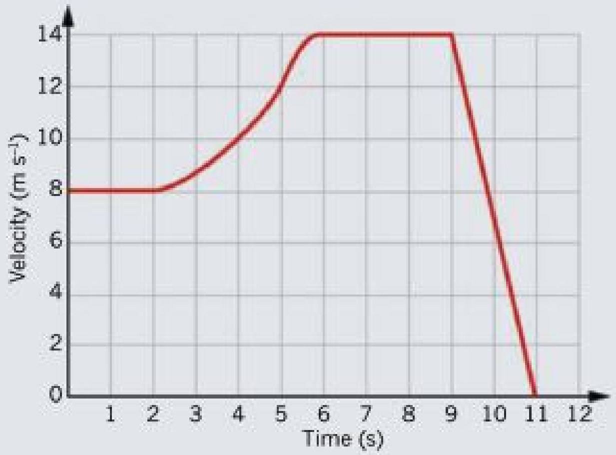




  a A car comes to a stop at a red light.
  b A swimmer is travelling at a constant speed.
  c A motorbike starts from rest with uniform acceleration.

Chapter review *continued*

10 This velocity–time graph is for an Olympic road cyclist as he travels, initially north, along a straight section of track.



a Calculate the displacement of the cyclist during his journey.
b Calculate the magnitude, to three significant figures, of the average velocity of the cyclist during this 11.0 s interval.
c Calculate the acceleration of the cyclist at $t = 1$ s.
d Calculate the acceleration of the cyclist at $t = 10$ s.
e Which one or more of the following statements correctly describes the motion of the cyclist?
  A He is always travelling north.
  B He travels south during the final 2 s.
  C He is stationary at $t = 8$ s
  D He returns to the starting point after 11 s.

11 A car starts from rest and has a constant acceleration of 3.5 m s^{-2} for 4.5 s. What is its final speed?

12 A jet-ski starts from rest and accelerates uniformly. If it travels 2.0 m in its first second of motion, calculate:
a its acceleration
b its speed at the end of the first second
c the distance the jet-ski travels in its second second of motion.

13 A skater is travelling along a horizontal skate rink at a speed of 10 m s^{-1}. After falling over, the skater takes 10 m to come to rest. Calculate, to two significant figures, the answers to the following questions about the skater's movement.
a What is the average acceleration of the skater?
b How long does it take the skater to come to a stop?

14 The graph shows the position of Candice dancing across a stage.

a What is Candice's starting position?
b In which of the sections (A–D) is Candice at rest?
c In which of the sections (A–D) is Candice moving in a positive direction, and what is her velocity?
d In which of the sections (A–D) is Candice moving with a negative velocity and what is the magnitude of this velocity?
e Calculate Candice's average speed during the 25 s motion.

15 A slingshot is used to launch a marble vertically into the air at 39.2 m s^{-1}. Discuss the velocity and acceleration of the marble as it travels to its maximum height. Indicate the time that it takes to reach the top. Consider 'up' as positive.

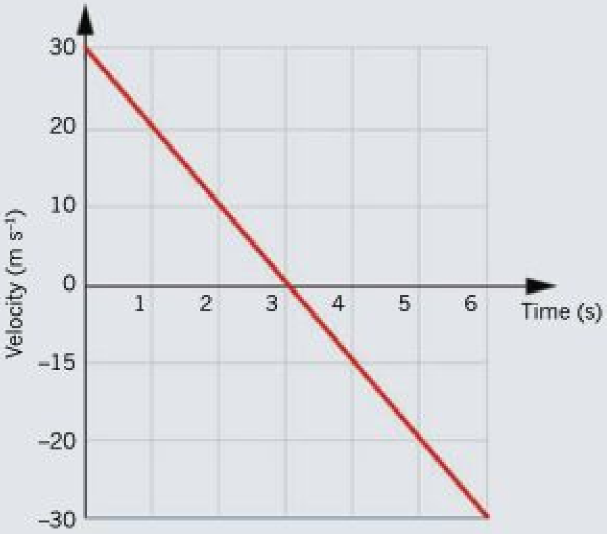
16 A golfer mis-hits a golf ball straight up into the air. Which one of the following statements best describes the acceleration of the ball while it is in the air?
A The acceleration of the ball decreases as it travels upwards, becoming zero as it reaches its highest point.
B The acceleration is constant as the ball travels upwards, then reverses direction and falls down again.
C The acceleration of the ball is greatest when the ball is at the highest point.

17 Steph tosses a rock vertically into the air. Which of the options below correctly fills the blanks of the following statement about the rock's motion?

On its way upwards, the rock has ________________ velocity and ________________ acceleration. At the highest point, the rock has ________________ velocity and ________________ acceleration. On its way downwards, the rock has ________________ velocity and ________________ acceleration.

A upwards; upwards; zero; downwards; downwards; downwards
B upwards; downwards; zero; downwards; downwards; downwards
C upwards; upwards; zero; zero; downwards; downwards
D upwards; downwards; zero; zero; downwards; downwards

18 After winning a tennis match, Claire hit a tennis ball vertically into the air at 30 m s^{-1}. The *v–t* and *a–t* graphs for the tennis ball are shown below. Use the graphs or the equations for uniform acceleration to answer the following questions. Use g = 10 m s^{-2} for these questions. Assume the motion in question is symmetrical, starting and ending at the same point.



- **a** What is the maximum height reached by the ball?
- **b** What is the time that the ball takes to return to its starting position?
- **c** What is the velocity of the ball 5.0 s after Claire hits it?
- **d** What is the acceleration of the ball at its maximum height?

19 A hot-air balloon is 80 m above the ground and travelling vertically downwards at 8.0 m s^{-1} when one of the passengers, Olivia, accidentally drops a coin over the side.

- **a** How long does the balloon take to reach the ground?
- **b** What is the speed of the coin as it reaches the ground?
- **c** How long after the coin reaches the ground does the balloon touch down?

The following information relates to questions 20 and 21. During a game of minigolf, Renee putts a ball so that it hits an obstacle and travels straight up into the air, reaching its highest point after 1.5 s.

20 What was the initial velocity of the ball as it launched into the air?

21 Calculate the maximum height reached by the ball.

CHAPTER 10 Momentum and force

In the seventeenth century, Sir Isaac Newton published three laws that explain why objects in our universe move as they do. These laws became the foundation of a branch of physics called mechanics: the science of how and why objects move. They have become commonly known as Newton's three laws of motion.

Using Newton's laws, this chapter will describe the relationship between the forces acting on an object and its motion. It will also discuss the relationship between force, period of time and change in momentum (impulse).

Key knowledge

By the end of this chapter, you will have studied the physics of momentum and forces, and will be able to:

- apply concepts of momentum to linear motion: $p = mv$
- explain changes in momentum as being caused by a net force: $F_{\text{net}} = \frac{\Delta p}{\Delta t}$
- model the force due to gravity, F_g, as the force of gravity acting at the centre of mass of a body, $F_g = mg$, where g is the gravitational field strength (9.8 N kg^{-1} near the surface of Earth)
- model forces as vectors acting at the point of application (with magnitude and direction), labelling these forces using the convention 'force on A by B' or $F_{\text{on A by B}} = -F_{\text{on B by A}}$
- apply Newton's three laws of motion to a body on which forces act: $a = \frac{F_{\text{net}}}{m}$, $F_{\text{on A by B}} = -F_{\text{on B by A}}$
- apply the vector model of forces, including vector addition and components of forces, to readily observable forces including the force due to gravity, friction and reaction forces
- analyse impulse (momentum transfer) in an isolated system (for collisions between objects moving in a straight line): $I = \Delta p$
- investigate and analyse theoretically and practically momentum conservation in one dimension.

10.1 Newton's first law

The previous chapter developed the concepts and ideas needed to describe the motion of a moving body. In this chapter, rather than simply describe the motion, you will investigate the forces that cause the motion to occur.

FORCE

In simple terms, a **force** can be thought of as a push or a pull, but forces exist in a wide variety of situations in your life and are fundamental to the nature of matter and the structure of the universe. Consider each of the photographs in Figure 10.1.1. For each situation a force—a push or pull—is acting.

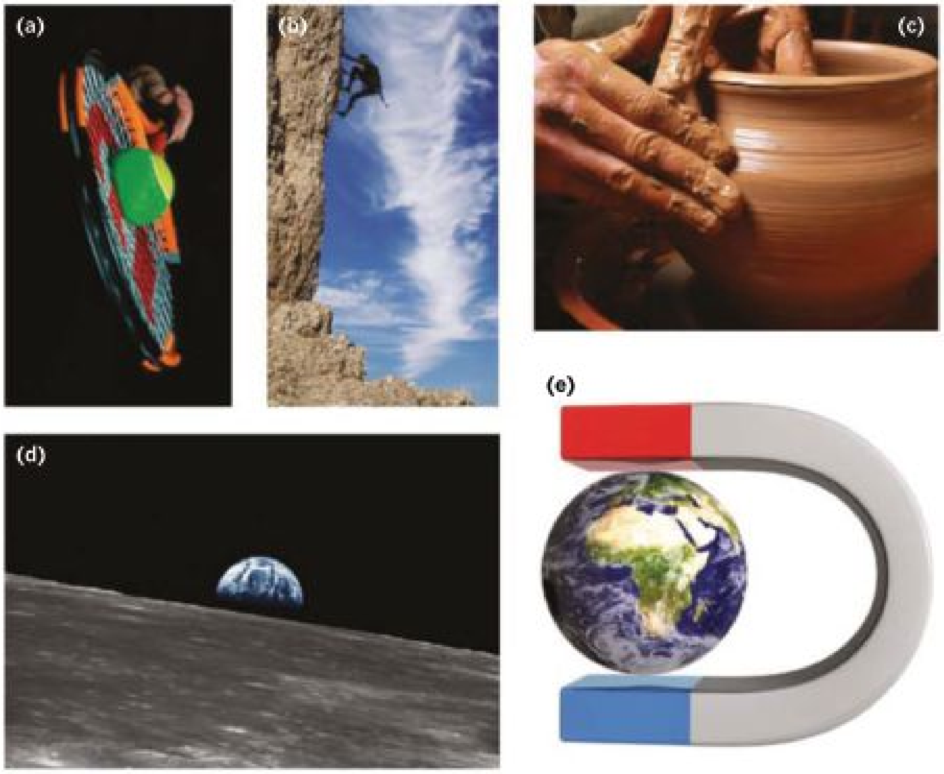


FIGURE 10.1.1 (a) At the moment of impact, both the tennis ball and the racquet strings are distorted by the forces acting at this instant. (b) The rock climber is relying on the frictional force between his hands and feet and the rock face. (c) A continual force causes the clay to deform into the required shape. (d) The gravitational force between the Earth and the Moon is responsible for two high tides each day. (e) The globe is suspended in mid-air because of the magnetic forces of repulsion and attraction.

In each of the situations depicted in Figure 10.1.1, forces are acting. Some are applied directly to an object and some act on a body without touching it. Forces that act directly on a body are called **contact forces**, because the body will only experience the force while contact is maintained. Forces that act on a body at a distance are **non-contact forces**.

Contact forces are the easiest to understand and include the simple pushes and pulls that are experienced daily in people's lives. Examples of these include the forces between colliding billiard balls and the forces that act between you and your chair as you sit reading this book. Friction and drag forces are also contact forces.

Non-contact forces occur when the object causing the push or pull is physically separated from the object that experiences the force. These forces are said to 'act at a distance'. Gravitation, magnetic and electric forces are examples of non-contact forces.

A force is a push or a pull which is measured in newtons (N). It is a vector and so it requires a magnitude and a direction to describe it fully.

The amount of force acting can be measured using the SI unit called the newton, which is given the symbol N. The unit, which will be defined later in the chapter, honours Sir Isaac Newton (1642–1727), who is still considered to be one of the most significant physicists to have lived and whose first law is the subject of this section. A force of one newton, 1 N, is approximately the force you have to exert when holding a 100 g mass against the downwards pull of gravity. In everyday life this is about the same as holding a small apple.

If more than one force acts on a body at the same time, the body behaves as if only one force—the vector sum of all the forces—is acting. The vector sum of the forces is called the resultant or **net force**, F_{net}. (Note: vectors were covered in detail in Chapter 8.)

The net force acting on a body experiencing a number of forces acting simultaneously is given by the vector sum of all the individual forces:

$$F_{net} = F_1 + F_2 + \ldots + F_n$$

NEWTON'S FIRST LAW

Inertia and Newton's first law are closely related; in fact, some people call Newton's first law the law of inertia. Inertia is the tendency of an object to maintain its velocity. This tendency is related to the mass of an object, so that the greater the mass, the harder it is to get it moving or to stop it from moving.

Newton's first law is a law that is often misunderstood due to a common misconception. People mistakenly think that an object that is moving at constant velocity must have a force causing it to move. This section will address this misconception and will enable you to understand how Newton's first law applies to all situations in which an object moves.

Newton's first law can be stated as:

An object will maintain a constant velocity unless an unbalanced, external force acts on it.

This statement needs to be analysed in more detail by first examining some of the key terms used. The term 'maintain a constant velocity' implies that, if the object is moving, then it will continue to move with a velocity that has the same magnitude and direction. For example, if a car is moving at 12.0 m s^{-1} south, then some time later it will still be moving at 12.0 m s^{-1} south (see Figure 10.1.2). It should also be noted that zero velocity can also be constant, so if the car is moving at 0 m s^{-1}, then some time later it will still be moving at 0 m s^{-1}.

FIGURE 10.1.2 A car maintaining a constant velocity.

The term 'unless' is particularly important in Newton's statement. In effect, it tells you what must *not* happen for the motion to be constant instead of telling you what must happen. Rephrasing Newton's first law to remove the term 'unless' means that it could say:

An object will not maintain its velocity if an unbalanced, external force is applied.

In this context, *not* continuing with its velocity implies that the object is changing its velocity. A change in velocity means that the object is accelerating.

PHYSICSFILE

The effects of forces

Applying a force can cause an object to speed up, to slow down, to start moving, to stop moving, or to change its direction. The effect depends on the direction of the force in relation to the direction of the velocity vector of the object experiencing the force. The effect of external forces is summarised in Table 10.1.1 below.

Relationship between velocity and force	Effect of force
force applied to object at rest	object starts moving
force in same direction as velocity	magnitude of velocity increases (object speeds up)
force in opposite direction to velocity	magnitude of velocity decreases (object slows down)
force perpendicular to velocity	direction of velocity changes (object turns)

TABLE 10.1.1 The effect of the application of a force, depending on the relationship between the direction of the force and the velocity.

In all cases, the effect of a force is to change the velocity of an object, whether it is the magnitude of the velocity, the direction of the velocity, or both that changes.

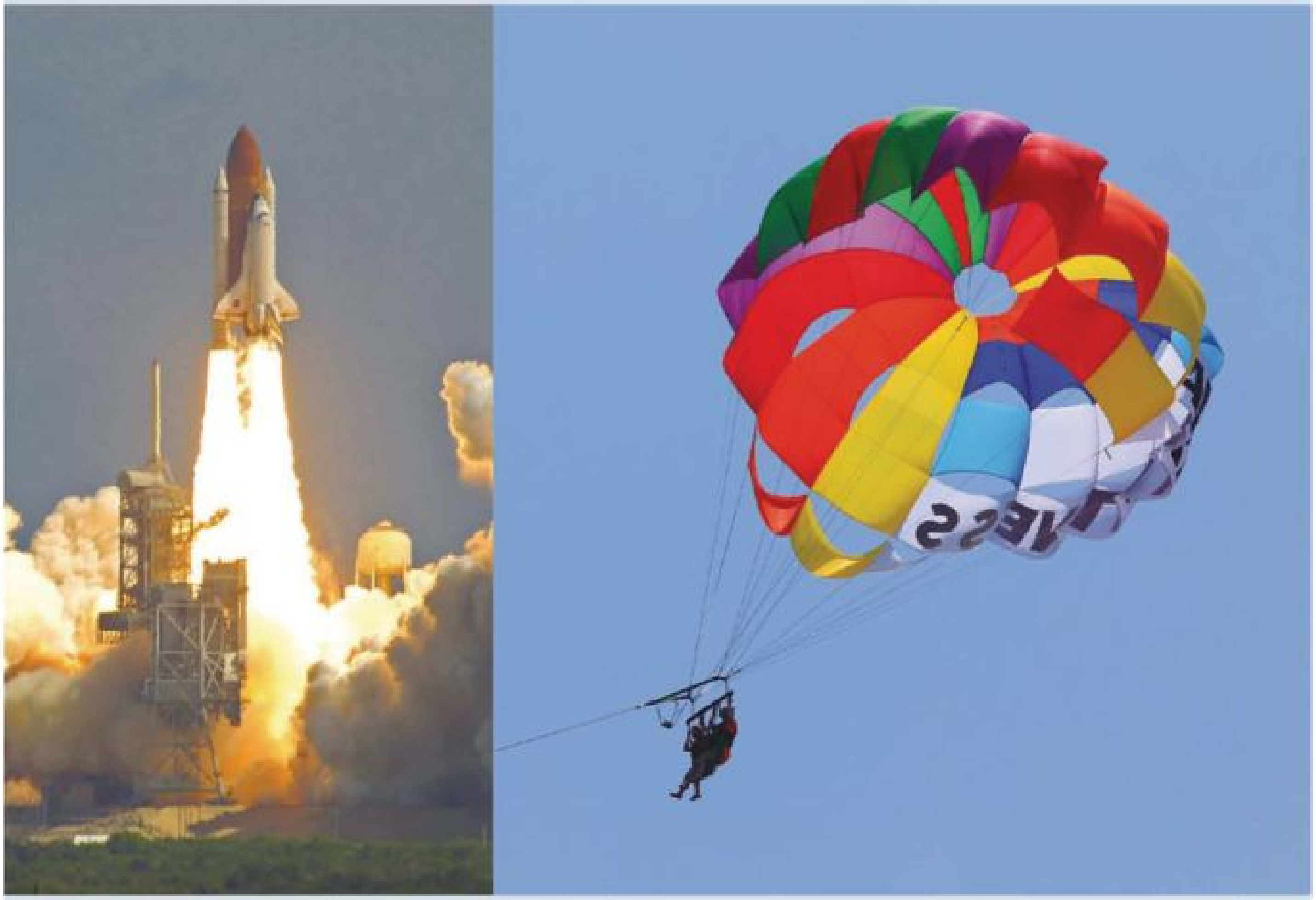


FIGURE 10.1.3 The upwards thrust force on a rocket has the effect of speeding it up, whereas the upwards drag force on a parachute has the effect of slowing the descent of a parachutist.

The use of the term 'unbalanced' in relation to the acting force implies that there must be a net force acting on the object. If the forces are balanced, then the object's velocity will remain constant. If the forces are unbalanced, then the velocity will change, or will not remain constant. This is illustrated in Figure 10.1.4.

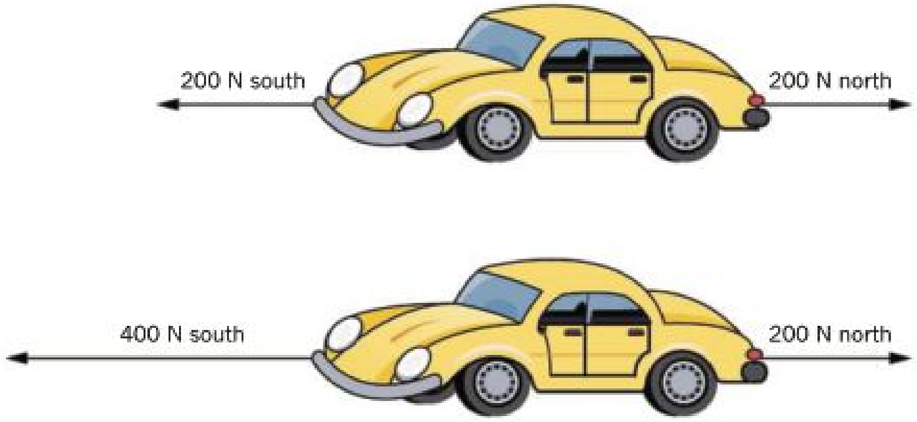


FIGURE 10.1.4 The forces on the top car are balanced, and so it will maintain a constant velocity. The forces on the bottom car are unbalanced: it has a net force in the forwards direction, so its velocity will change.

The term 'external', in relation to forces, implies that the forces are not internal. When forces are internal, they will have no effect on the motion of the object. For example, if you are sitting in a car and push forwards on the steering wheel then the car will not move forwards due to this force. In order for you to push forwards on the steering wheel, you must push backwards on the seat. Both the steering wheel and the seat are attached to the car, therefore there are two forces acting on the car that are equal and in the opposite direction to each other, as shown in Figure 10.1.5. All internal forces must result in balanced forces on the object and therefore they will not change the velocity of the object.

FIGURE 10.1.5 This driver is applying internal forces on a car. These internal forces will balance and cancel each other.

Stating Newton's first law in different ways

Ludwig Wittgenstein, an Austrian–British philosopher, suggested that 'understanding means seeing that the same thing said different ways is the same thing'. To truly understand Newton's first law, you should be able to state it in different ways yet still recognise it as being consistent with Newton's first law.

All of these statements are consistent with Newton's first law:

- An object will maintain a constant velocity unless an unbalanced, external force acts on it.
- An object will continue with its motion unless an unbalanced, external force is applied.
- An object will not continue with its velocity if an unbalanced, external force is applied.
- A body will either remain at rest or continue with constant speed in a straight line (i.e. constant velocity) unless it is acted on by a net force.

- If an unbalanced, external force is applied, then an object's velocity will change.
- If a net force is applied, then the object's velocity will change.
- If no net force is applied, the object will not accelerate.
- If a net force is applied, an acceleration will result.
- Net forces cause acceleration.
- No force, no acceleration.
- Constant velocity means no net force is applied.

INERTIA

Inertia is considered to be the resistance to a change in motion of an object. It is related to the mass of the object. As the mass of the object increases, the inertia increases and therefore:

- it becomes harder to start it moving if it is stationary, or
- it becomes harder to stop it moving, or
- it becomes harder to change its direction of motion.

You can experience the effect of inertia when you push a trolley in a supermarket. If the trolley is empty, it is relatively easy to start pushing it, or to pull it to a stop when it is already moving. It is also easy to turn a corner. If you fill the trolley with heavy groceries, you notice that it becomes more difficult—that is, it requires more force—to make the trolley start moving when it is at rest, and it becomes more difficult to pull it to a stop if it is already moving. It also requires more force to change the trolley's direction.

It is important to note that the effect of inertia is independent of gravity. Since inertia depends on mass, and weight force due to gravity also depends on mass, it is a common misunderstanding to think that the effects of inertia only apply in the presence of gravity. However, even in space it would be just as difficult to change the state of motion of the trolley, as described above, as it is in the supermarket aisle.

Newton's first law and inertia

The connection between Newton's first law and inertia is very close. Due to inertia, an object will continue with its motion unless a net force acts on the object.

You experience the connection between Newton's first law and inertia if you are standing on public transport. Imagine standing on a tram that is initially at rest and then starts moving forwards. If you are not holding on to anything, you may stumble backwards as though you have been pushed backwards. However, you have not been pushed backwards; the tram has started moving forwards and, since you have inertia, your mass resists the change in motion. According to Newton's first law, your body is simply maintaining its original state of being motionless until an unbalanced force acts to accelerate it. When the tram later comes to a sudden stop, your body again resists the change by continuing to move forwards until an unbalanced force acts to bring it to a stop.

OBSERVING NEWTON'S FIRST LAW

When an object is in motion—for example, a pen sliding across a table—it will eventually stop. It may not seem obvious, but this is a very good example of Newton's first law. The motion does not continue; therefore a net force must be acting. In this case, however, the force is not an obvious one. Confusion sometimes arises if the force due to friction is overlooked. Friction is a force that always acts in the opposite direction to the motion of objects. Air resistance is also a force that is often overlooked, as is the force due to gravity. By ignoring the effect of these important forces, it can be easy to come to the incorrect conclusion that the natural state of any object is to be at rest. By considering all the external forces acting on an object, it becomes clear that the natural state of any object is to maintain whatever velocity it currently has.

PHYSICS IN ACTION

Terminal velocity

In the previous chapter it was stated that, in the absence of air resistance, all objects accelerate towards the surface of Earth at a constant rate of 9.8 m s^{-2}. However, air resistance increases as the speed of an object increases, and it becomes very significant when objects are moving at high speeds. Newton's first law can be used to explain how air resistance causes a skydiver to experience terminal velocity. As the skydiver begins falling towards Earth the only external force is the weight force. The weight force causes the skydiver to accelerate at 9.8 m s^{-2}. As the skydiver gets faster, air resistance pushes upwards. This reduces the magnitude of the net force, which decreases the skydiver's acceleration. Eventually the air resistance becomes so great that it exactly balances the weight force. According to Newton's first law, the skydiver maintains a constant velocity when all the external forces are balanced: this velocity is terminal velocity. Terminal velocity will be different for different objects because air resistance depends on the object's shape. For example, terminal velocity for a sheet of paper is much less than for a skydiver.

FIGURE 10.1.6 When a skydiver reaches terminal velocity, weight force is in balance with air resistance.

EXTENSION

Frictional forces

Friction is a force that opposes movement. Suppose you want to push your textbook along the table. As you start to push the book, you find that the book does not move at first. You then increase the force that you apply. Suddenly, at a certain critical value, the book starts to move.

There is a maximum frictional force that resists the start of the slide. This force is called the static friction force, F_s. Once the book begins to slide, a much lower force than F_s is needed to keep the book moving. This force is called the kinetic friction force and is represented by F_k. The graph in Figure 10.1.7 shows how the force required to move an object changes as static friction is overcome.

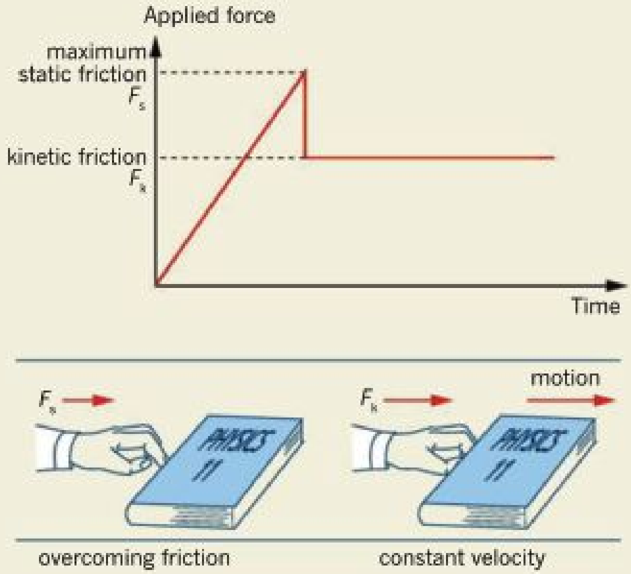


FIGURE 10.1.7 To get things moving, the static friction between an object and the surface must be overcome. This requires a larger force than is needed to maintain constant velocity.

This phenomenon can be understood when you consider that even the smoothest surfaces are quite jagged at the microscopic level. When the book is resting on the table, the jagged points of its bottom surface have settled into the valleys of the surface of the table and this helps to resist attempts to slide the book. Once the book is moving, the surfaces do not have any time to settle into each other, so less force is required to keep the book moving.

Another fact that helps to explain friction arises from the forces of attraction between the atoms and molecules of the two different surfaces that are in contact. These surfaces produce weak bonding between their respective particles, so before one surface can move across the other, these bonds must be broken. This extra effort adds to the static friction force. Once there is relative motion between the surfaces, the bonds cannot re-form.

In everyday life, there are situations in which friction is desirable (e.g. walking) and others in which it is a definite problem. Consider the moving parts within the engine of a car. Friction can rob an engine of its fuel economy and cause it to wear out. Special oils and lubricants are introduced in order to prevent moving metal surfaces from touching. If the moving surfaces actually moved over each other, they would quickly wear, producing metal filings that could damage the engine. Instead, metal surfaces are separated by a thin layer of oil. The oils are chosen on the basis of their viscosity (thickness). For example, low viscosity oils can be used in the engine, while heavier oils are needed in the gearbox and differential of the car where greater forces are applied to the moving parts. On the other hand, the lack of friction enables magnetic levitation trains to reach very high speeds, as shown in Figure 10.1.8.

FIGURE 10.1.8 This magnetic levitation train in China rides 1 cm above the track, so the frictional forces are negligible. The train is propelled by a magnetic force to a cruising speed of about 430 km h^{-1}.

At other times, having friction is essential. If there is any ice on the road when driving to snowfields, drivers are required to fit chains to their cars. When driving over a patch of ice, the chains break through the ice and the car is able to grip the road. Similarly, friction is definitely required within the car's brakes when the driver wants to slow down. In fact, modern brake pads are specially designed to maximise the friction between them and the brake drum or disc.

PHYSICS IN ACTION

Galileo's law of inertia

Galileo Galilei (Figure 10.1.9) was born into an academic family in Pisa, Italy, in 1564. He made significant contributions to physics, mathematics and scientific method through intellectual rigour and the quality of his experimental design. But, more than this, Galileo helped to change the way the universe is viewed.

FIGURE 10.1.9 Galileo Galilei.

Galileo's most significant contributions were in astronomy. Through his development of the refracting telescope he discovered sunspots, lunar mountains and valleys, the four largest moons of Jupiter (now called the Galilean Moons, one of which is shown in Figure 10.1.10) and the phases of Venus. In mechanics, he demonstrated that projectiles moved with a parabolic path and that different masses fall at the same rate (the law of falling bodies).

FIGURE 10.1.10 The major moons of Jupiter, including Io, are known as the Galilean Moons.

These developments were important because they changed the framework within which mechanics was understood. This framework had been in place since Aristotle had constructed it in the fourth century BCE. Aristotle's thesis was based on the observation that a moving body's natural state is at rest, and the object will come to rest unless a force is applied. However, Galileo's experiments led him to believe that the natural state of an object is not at rest. He suggested that objects maintained their state of motion. He called this tendency inertia.

By the sixteenth century, the work of the Greek philosophers had become entrenched, being widely supported in the universities and at a political level. In Italy at that time, government was controlled by the Roman Catholic Church. One might think that Galileo would have won praise from his peers for making such progress, but so ingrained and supported was the Aristotelian view that Galileo actually lost his job as a professor of mathematics in Pisa in 1592.

Galileo was not, however, without supporters, and he was able to move from Pisa to Padua, where he continued teaching mathematics. In Padua, Galileo began to use measurements from carefully constructed experiments to strengthen his ideas. He entered into vigorous debate in which his ideas (founded on observation) were pitted against the philosophy of the past and the politics of the day. The most divisive debate involved the motion of the planets. The ancient Greek view, formalised by Ptolemy in the second century CE, was that Earth was at the centre of the solar system and that the planets, the Moon and the Sun were in orbit around it. This view was taught by the Church and was also supported by common sense.

In 1630, Galileo published a book in which he debated the Ptolemaic view and the new Sun-centred model proposed by Copernicus. On the basis of his own observations, Galileo supported the Copernican view of the universe. Although the book had been passed by the censors of the day, Galileo was summoned to Rome to face the Inquisition for heresy (opposition to the Church). The finding went against Galileo and all copies of his book had to be burned. He was sentenced to permanent house arrest for the rest of his life.

Galileo died in 1642 in a village near Florence. He had become an influential thinker across Europe, and the scientific revolution he had helped to start accelerated in the freer Protestant countries in northern Europe. For its part, the Roman Catholic Church under Pope John Paul II in 1979 began an investigation into Galileo's trial, and in 1992 a papal commission reversed the Church's condemnation of him.

10.1 Review

SUMMARY

- A force is a push or a pull. Some forces act on contact while others can act at a distance.
- Force is a vector quantity whose SI unit is the newton (N).
- Newton's first law can be written in many ways:
  - An object will continue with its velocity unless an unbalanced force causes the velocity to change.
  - Net forces cause acceleration.
- Inertia is the tendency of an object to resist changes in motion.
- Inertia is related to mass; an object with a large mass will have a large inertia.

KEY QUESTIONS

1 A student observes a box sliding across a surface and slowing down to a stop. From this observation what can the student conclude about the forces acting on the box?

2 A car changes its direction as it turns a bend in the road while maintaining its speed of 16 m s^{-1}. From this, what can you conclude?

3 A bowling ball rolls along a smooth wooden floor at constant velocity. Ignoring the effects of friction and air resistance, which of the following options, relating to the force acting on the ball, is correct?

A There must be a net force acting forwards to maintain the velocity of the ball.

B There must not be an unbalanced force acting on the ball.

C The forwards force acting on the ball must be balanced by the friction that opposes the motion.

D More information is needed.

4 If a person is standing up in a moving bus that stops suddenly, the person will tend to fall forwards. Has a force acted to push the person forwards? Use Newton's first law to explain what is happening.

5 What horizontal force has to be applied to a wheelie bin if it is to be wheeled to the street on a horizontal path against a frictional force of 20 N at a constant 1.5 m s^{-1}?

6 A young boy is using a horizontal rope to pull his billycart at a constant velocity. A frictional force of 25 N also acts on the billycart.

a What force must the boy apply to the rope?

b The boy's father then attaches a longer rope to the cart because the short rope is uncomfortable to use. The rope now makes an angle of 30° to the horizontal. What is the horizontal component of the force that the boy needs to apply in order to move the cart with constant velocity?

c What is the tension force acting along the rope that the boy must supply?

7 Passengers on commercial flights are required to be seated and have their seatbelts done up when their plane is coming in to land. What would happen to a person who was standing in the aisle as the plane travelled along the runway during landing?

8 Consider the following situations, and name the force that causes each object to travel along a path which is not a straight line.

a The Earth moves in a circle around the Sun with constant speed.

b An electron orbits the nucleus with constant speed.

c A cyclist turns a corner at constant speed.

d An athlete swings a hammer in a circle with constant speed.

9 A magician performs a trick in which a cloth is pulled quickly from under a glass filled with water without causing the glass to fall over or the water to spill out.

a Explain the physics principles underlying this trick.

b Does using a full glass make the trick easier or more difficult? Explain.

10 Which of these objects would find it most difficult to come to a stop: a cyclist travelling at 50 km h^{-1}, a car travelling at 50 km h^{-1} or a fully laden semitrailer travelling at 50 km h^{-1}? Explain your answer.

11 When flying at constant speed at a constant altitude, a light aircraft has a weight of 50 kN down, and the thrust produced by its engines is 12 kN to the east. What is the lift force required by the wings of the plane, and how large is the drag force that is acting?

10.2 Newton's second law

Newton's second law makes the quantitative connection between force, mass and acceleration.

Newton's second law helps to resolve the misconception that many people have about the time taken for objects of different mass to fall to the ground. Many mistakenly believe that heavy objects will fall faster than lighter objects. Once again, air resistance acts to complicate the matter and results in the observation of different times for different masses to fall the same distance. However, even when air is removed, the misconception that larger masses fall faster than lighter masses still persists.

Figure 10.2.1 depicts a famous experiment, mentioned in Section 9.5 of this text, which shows that objects fall together when the effect of air resistance is removed. A web search for 'hammer and feather on the Moon' will enable you to view a video of David Scott's 1971 experiment. Although the images are quite poor, you should be able to see both objects accelerating at the same rate.

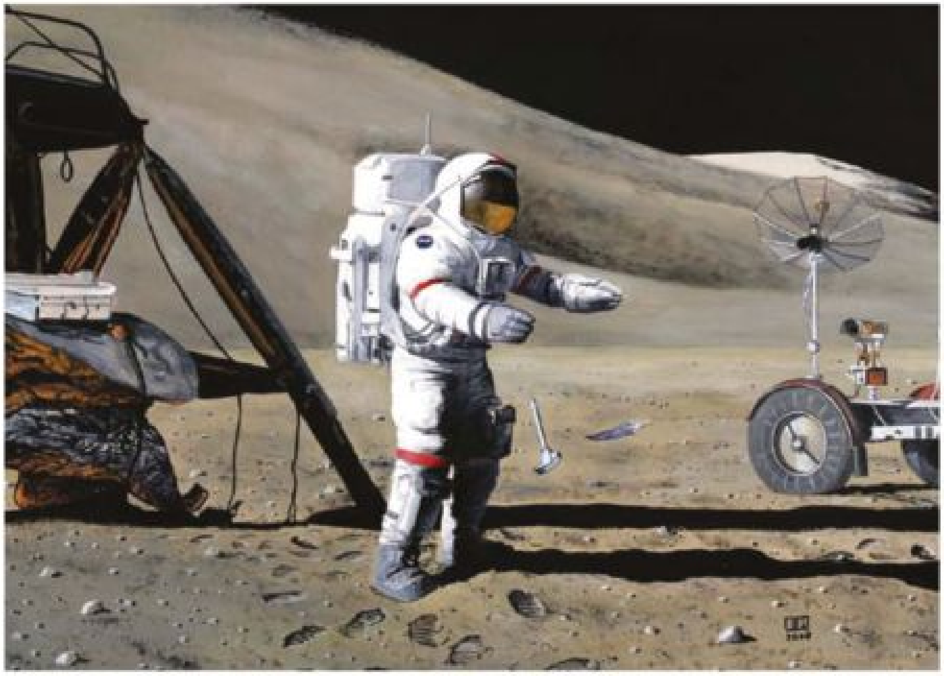


FIGURE 10.2.1 An artist's image of the famous hammer and feather experiment conducted on the Moon.

NEWTON'S SECOND LAW

Newton's second law of motion states that:

> The acceleration of an object is directly proportional to the net force on the object and inversely proportional to the mass of the object:
>
> $a = \frac{F_{net}}{m}$
>
> where a is the acceleration of an object (in m s^{-2})
>
> F_{net} is the force applied to the object (in N)
>
> m is the mass of the object (in kg).

The above equation is also commonly written as $F_{net} = ma$.

By definition, 1 **newton** is the force needed to accelerate a mass of 1 kg at 1 m s^{-2}.

The unit of force is the units of mass and acceleration combined, or kg m s^{-2}. This unit was renamed the newton (N) in honour of Sir Isaac Newton.

One of the implications of Newton's second law is that, for a given mass, a greater acceleration is achieved by applying a greater force. This is shown in Figure 10.2.2. Doubling the applied force will double the acceleration of the object. In other words, acceleration is proportional to the net force applied.

Notice also in Figure 10.2.2 that the acceleration of the object is in the same direction as the net force applied to it.

Newton's second law also explains how acceleration is affected by the mass of an object. For a given force, the acceleration of an object will decrease with increased mass. In other words, acceleration is inversely proportional to the mass of an object. This is shown in Figure 10.2.3.

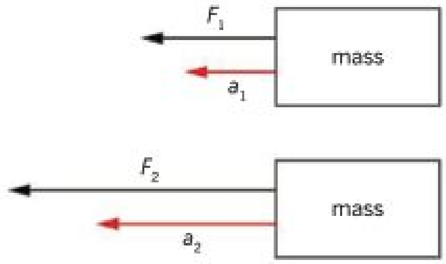


FIGURE 10.2.2 Given the same mass, a larger force will result in a larger acceleration. If the force is doubled, then the acceleration is also doubled.

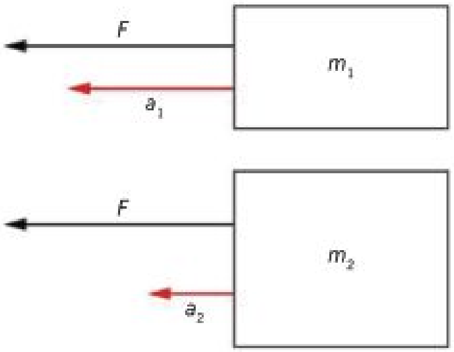


FIGURE 10.2.3 Given the same force, a larger mass will result in a lower acceleration. If the mass is doubled, then the acceleration is halved.

PHYSICSFILE

Maximising acceleration

Dragster race cars are designed to achieve the maximum possible acceleration in order to win a race in a straight line over a relatively short distance. According to Newton's second law, acceleration is increased by increasing the applied force and by reducing the mass of the object. For this reason, dragster race cars are designed with very powerful engines that produce an enormous forwards force and an aerodynamic shape to minimise air resistance. There is not much else to the car, so this helps to minimise the mass.

Newton's second law also helps you understand why a motorcycle can accelerate from the lights at a greater rate than a car or a truck. While the engines in a car or truck are usually more powerful than a motorcycle engine, the motorcycle has much less mass, which allows for greater acceleration.

FIGURE 10.2.4 The aerodynamic design of motorcycles and their lower mass enables them to accelerate faster than cars and trucks.

Observing Newton's second law

The equation $F_{net} = ma$ enables you to calculate the force that causes a mass to accelerate. Mass is something that is easily experienced. You can measure the mass of an object on a balance. You can hold two masses in your hands and feel their effect. Similarly, you can observe acceleration, for example when a car accelerates away from the traffic lights. Force, on the other hand, is not something you can see. However, you can see the *effect* of a force.

Worked example 10.2.1

CALCULATING THE FORCE THAT CAUSES AN ACCELERATION

Calculate the net force causing a 5.50 kg mass to accelerate at 3.75 m s^{-2} west.	
Thinking	**Working**
Ensure that the variables are in their standard units.	$m = 5.50$ kg $a = 3.75$ m s^{-2} west
Apply the equation for force from Newton's second law.	$F_{net} = ma$ $= 5.50 \times 3.75$ $= 20.6$ N
Give the direction of the net force, which is always the same as the direction of the acceleration.	$F_{net} = 20.6$ N west

Worked example: Try yourself 10.2.1

CALCULATING THE FORCE THAT CAUSES AN ACCELERATION

Calculate the net force causing a 75.8 kg runner to accelerate at 4.05 m s^{-2} south.

The first equation for uniform acceleration, which is discussed in Section 9.4 on page 315, can be combined with Newton's second law to calculate changes in time or velocity.

The first equation for uniform acceleration is:

$v = u + at$

This can be rearranged to give:

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

Combining this with $F_{net} = ma$ gives:

$F_{net} = m\left(\frac{v - u}{t}\right)$

Worked example 10.2.2

CALCULATING THE FINAL VELOCITY OF AN ACCELERATING MASS

Calculate the final velocity of a 225 kg scooter that accelerates for 2.00 s from rest due to a force of 2430 N north.	
Thinking	**Working**
Ensure that the variables are in their standard units.	$m = 225$ kg $t = 2.00$ s $u = 0$ m s^{-1} $F_{net} = 2430$ N north
Apply a variation of the equation for force from Newton's second law.	$F_{net} = m\frac{(v - u)}{t}$ $(v - u) = \frac{F_{net}t}{m}$ $v = \frac{F_{net}t}{m} + u$ $= \frac{2430 \times 2.00}{225} + 0$ $= 21.6$ m s^{-1}
Give the direction of the final velocity as being the same as the direction of the force.	$v = 21.6$ m s^{-1} north

Worked example: Try yourself 10.2.2

CALCULATING THE FINAL VELOCITY OF AN ACCELERATING MASS

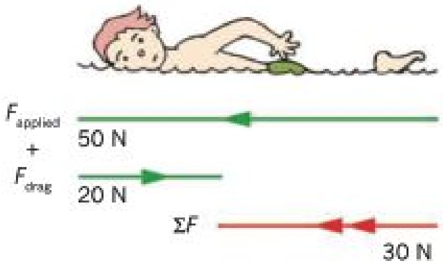
Calculate the final velocity of a 307 g fish that accelerates for 5.20 seconds from rest due to a force of 0.250 N left.

Forces do not always act alone. Mostly, more than one force will act on an object at any time. The overall effect of the forces depends on the direction of each of the forces. For example, some forces act together and some may oppose each other. When using Newton's second law, it is important to use the net, or resultant, force in the calculation. As forces are vectors, they can be added or combined using the techniques discussed in Chapter 8. Consider the following worked examples.

Worked example 10.2.3

CALCULATING THE ACCELERATION OF AN OBJECT WITH MORE THAN ONE FORCE ACTING ON IT

A swimmer whose mass is 75 kg applies a force of 50 N as he starts a lap. The water opposes his efforts to accelerate with a drag force of 20 N. What is his initial acceleration?

Thinking	**Working**
Determine the individual forces acting on the swimmer, and apply the vector sign convention.	F_1 = 50 N forwards = 50 N F_2 = 20 N backwards = –20 N
Determine the net force acting on the swimmer.	$F_{net} = F_1 + F_2$ = 50 + (–20) = +30 N or 30 N forwards 
Use Newton's second law to determine acceleration.	$a = \frac{F_{net}}{m}$ $= \frac{30}{75}$ = 0.40 m s^{-2} forwards

Worked example: Try yourself 10.2.3

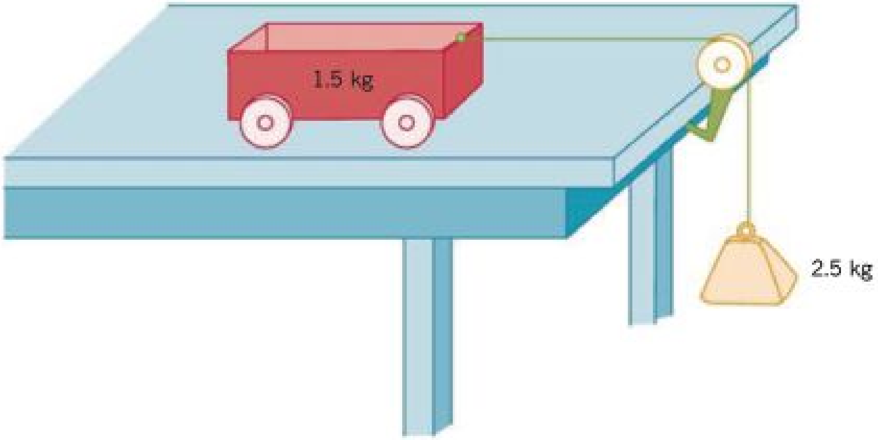
CALCULATING THE ACCELERATION OF AN OBJECT WITH MORE THAN ONE FORCE ACTING ON IT

A car with a mass of 900 kg applies a driving force of 3000 N as it starts moving. Friction and air resistance oppose the motion of the car with a force of 750 N. What is the car's initial acceleration?

Worked example 10.2.4

CALCULATING THE ACCELERATION OF A CONNECTED BODY

A 1.5 kg trolley cart is connected by a cord to a 2.5 kg mass as shown. The cord is placed over a pulley and allowed to fall under the influence of gravity.



a Assuming that the cart can move over the table unhindered by friction, determine the acceleration of the cart.

Thinking	**Working**
Recognise that the cart and the falling mass are connected, and determine a sign convention for the motion.	As the mass falls, the cart will move to the right. Therefore, both downwards movement of the mass and rightwards movement of the cart will be considered positive motion.
Write down the data that is given. Apply the sign convention to vectors.	$m_1 = 2.5$ kg $m_2 = 1.5$ kg $g = 9.8$ m s^{-2} down $= +9.8$ m s^{-2}
Determine the forces acting on the system.	The only force acting on the combined system of the cart and mass is the weight of the falling mass. $F_{\text{net}} = F_g$ $= m_1 g$ $= 2.5 \times 9.8$ $= 24.5$ N in the positive direction
Calculate the mass being accelerated.	This net force has to accelerate not only the cart but also the falling mass. $m_1 + m_2 = 2.5 + 1.5$ $= 4.0$ kg
Use Newton's second law to determine acceleration.	$a = \frac{F_{\text{net}}}{m}$ $= \frac{24.5}{4.0}$ $= 6.1$ m s^{-2} to the right

b If a frictional force of 8.5 N acts against the cart, what is the acceleration now?

Thinking	Working
Write down the data that is given. Apply the sign convention to vectors.	$m_1 = 2.5$ kg $m_2 = 1.5$ kg $g = 9.8$ m s^{-2} down $= +9.8$ m s^{-2} $F_{fr} = 8.5$ N left $= -8.5$ N
Determine the forces acting on the system.	There are now two forces acting on the combined system of the cart and mass: the weight of the falling mass and friction. $F_{net} = F_g + F_{fr}$ $= 24.5 + (-8.5)$ $= 16.0$ N $= 16.0$ N in the positive direction
Use Newton's second law to determine acceleration.	$a = \frac{F_{net}}{m}$ $= \frac{16.0}{4.0}$ $= 4.0$ m s^{-2} to the right

Worked example: Try yourself 10.2.4

CALCULATING THE ACCELERATION OF A CONNECTED BODY

A 0.6 kg trolley cart is connected by a cord to a 1.5 kg mass. The cord is placed over a pulley and allowed to fall under the influence of gravity.

a Assuming that the cart can move over the table unhindered by friction, determine the acceleration of the cart.

b If a frictional force of 4.2 N acts against the cart, what is the acceleration now?

THE FEATHER AND HAMMER EXPERIMENT

Why the experiment works on the Moon

When two objects with different mass fall under the influence of the force due to gravity in the absence of air resistance, they will both fall at the same rate. That is, their accelerations will be the same. They will cover the same displacement in the same time and will hit the ground at the same time if dropped from the same height. This experiment works on the Moon because there is no atmosphere and, therefore, no air resistance.

If you understand that all objects accelerate due to gravity at the same rate in a vacuum, the common misconception is that the force due to gravity is the same on all objects. This is not true. In fact, the force due to gravity is larger on larger objects and smaller on smaller objects. You mustn't forget that the objects have different masses. A larger mass experiences a greater force due to gravity (weight) than a smaller mass, but it also has more inertia so it requires that greater force in order to achieve the same acceleration. Refer to Figure 10.2.5 to see how this works.

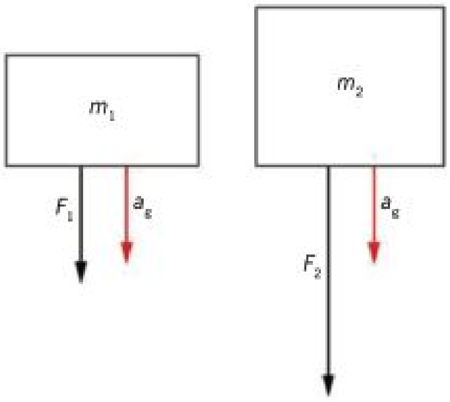


FIGURE 10.2.5 Given the same acceleration, a larger mass must have a larger force acting on it. If the mass is doubled, then the force is doubled.

If you consider the relationship between acceleration and mass, and the relationship between acceleration and force, then:

$$a = \frac{F}{m}$$

If m_2 is ten times the mass of m_1, then the force due to gravity on m_2 is ten times the force due to gravity on m_1. Consider the acceleration of both masses:

$$a_1 = \frac{F_1}{m_1}$$

If $F_2 = 10F_1$ and $m_2 = 10m_1$, then:

$$\begin{aligned} a_2 &= \frac{F_2}{m_2} \\ &= \frac{10F_1}{10m_1} \\ &= \frac{F_1}{m_1} \\ &= a_1 \end{aligned}$$

This proof shows that the ratio of force to mass is equal for all combinations of force and mass under the same effects due to gravity. This proves that all masses will experience the same acceleration if air resistance is removed.

Why the experiment does not work on Earth

When you see a feather floating down through the air, you know that it is accelerating at a rate far less than a hammer falling from the same height. From the previous section, you will know that the hammer and the feather have forces due to gravity acting on them that are proportional to their mass. They do not fall at the same rate because of the force of air resistance. Remember, Newton's second law says the acceleration is proportional to the *net* force acting on an object, which means you must consider all the forces acting on an object to determine the acceleration.

Air resistance is a force that results from air molecules colliding with the object. The faster the object moves, the greater the air resistance. In addition, the greater the surface area perpendicular to the direction of motion, the greater the air resistance is. This force, which acts in the opposite direction to the motion of an object, is significant when compared with the weight of the feather, but insignificant when compared with the weight of the hammer. As a result, this force has a noticeable effect on the feather's acceleration but makes no noticeable difference to the hammer's acceleration.

In Figure 10.2.6, the force of air resistance is denoted as F_{AR} and is the same size on both objects. The difference between the two objects is the downwards weight force (due to gravity).

Figure 10.2.6 also shows that the acceleration of the thinner object is much less than the acceleration of the thicker object. This is the observation that often causes misconceptions.

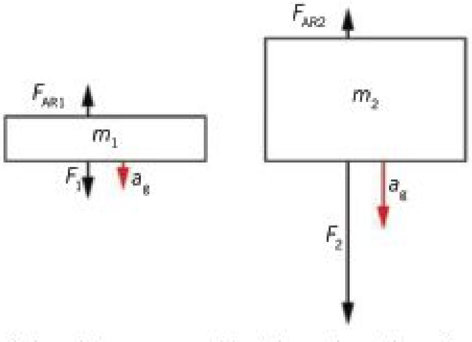


FIGURE 10.2.6 The effect of air resistance on an object depends on the surface area perpendicular to the motion and the speed of the object.

10.2 Review

SUMMARY

- Newton's second law states: The acceleration of an object is directly proportional to the net force on the object and inversely proportional to the mass of the object.
- Force can be calculated using the following formulas:

  $F_{net} = ma$
- This can be rewritten as:

  $F_{net} = m\left(\frac{v - u}{t}\right)$
- Force is difficult to perceive when it acts on objects, but we can perceive mass and acceleration.
- Different forces due to gravity act on different masses to cause the same acceleration.
- Air resistance is a force that acts to decrease the acceleration of objects moving through air.

KEY QUESTIONS

Use $g = 9.8$ m s^{-2} when answering the following questions.

1 Calculate the acceleration of a 23.9 kg mass when a net force of 158 N north acts on it.

2 Calculate the mass of an object if it accelerates at 9.20 m s^{-2} east when a net force of 352 N east acts on it.

3 Calculate the final velocity of a stationary 55.9 kg mass when a net force of 56.8 N north acts on it for 3.50 seconds.

4 Calculate the acceleration of a 45.0 kg mass that has a net force of 441 N acting on it due to gravity.

5 Calculate the acceleration of a 90.0 kg mass that has a net force of 882 N acting on it due to gravity.

6 Calculate the final velocity of a 60.0 kg mass moving at 2.67 m s^{-1} east, when a net force of 45.5 N west acts on it for 2.80 seconds.

7 Calculate the acceleration of a 60.9 g golf ball when a net force of 95.0 N south acts on it.

8 Calculate the mass of a train if it accelerates at 7.20 m s^{-2} north when a net force of 565 000 N north acts on it. Give your answer to three significant figures.

9 Calculate the final velocity of a stationary 3.00 g marble when a net force of 0.0823 N north acts on it for 0.0105 seconds.

10 Mary is paddling a canoe. The paddles are providing a constant driving force of 45 N south and the drag forces total 25 N north. The mass of the canoe is 15 kg and Mary has a mass of 50 kg.

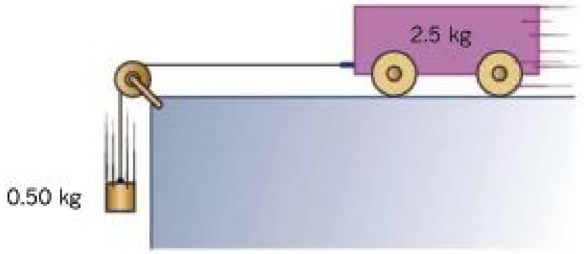
a What is Mary's mass?

b Calculate Mary's weight.

c Find the net horizontal force acting on the canoe.

d Calculate the magnitude of the acceleration of the canoe.

11 A 0.50 kg metal block is attached by a piece of string to a dynamics cart, as shown below. The block is allowed to fall from rest, dragging the cart along. The mass of the cart is 2.5 kg.



a If friction is ignored, what is the acceleration of the block as it falls?

b How fast will the block be travelling after 0.50 s?

c If a frictional force of 4.3 N acts on the cart, what is its acceleration?

12 An empty truck of mass 2000 kg has a top acceleration of 2.0 m s^{-2}. The mass of one box is 300 kg. How many boxes would be loaded if the truck's top acceleration decreased to 1.25 m s^{-2}.

13 The thrust force of a rocket with a mass of 50 000 kg is 1 000 000 N. Calculate its acceleration.

10.3 Newton's third law

FIGURE 10.3.1 A hammer hitting a nail is a good example of an action–reaction pair and Newton's third law.

Newton's first two laws of motion describe the motion of an object resulting from the forces that act on that object. Newton's third law of action and reaction is easily stated and seems to be widely known by students, but it is often misunderstood and misused. It is a very important law in physics as it assists with the understanding of the origin and nature of forces. Newton's third law is explored in detail in this section.

NEWTON'S THIRD LAW

Newton realised that all forces exist in pairs and that each force in the pair acts on a different object. Look at Figure 10.3.1, which shows a hammer hitting a nail on the head. Both the hammer and the nail experience forces during this interaction. The nail experiences a downwards force as the hammer hits it. When the nail is hit it moves a distance into the wood. As it hits the nail, the hammer experiences an upwards force that causes the hammer to stop. These forces are known as an action–reaction pair and are shown in Figure 10.3.2.

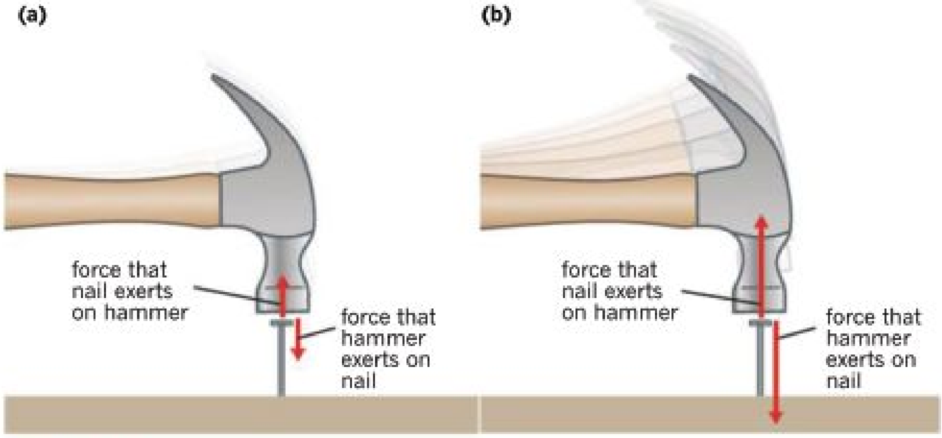


FIGURE 10.3.2 (a) As the hammer gently taps the nail, both the hammer and the nail experience small forces. (b) When the hammer smashes into the nail, both the hammer and the nail experience large forces. In both cases these forces are designated by $F_{\text{on nail by hammer}}$ and $F_{\text{on hammer by nail}}$.

It is important to note that, regardless of whether the hammer exerts a small or a large force on the nail, the nail will exert exactly the same-size force back on the hammer.

Newton's third law states:

For every action (force), there is an equal and opposite reaction (force).

This means that when object A exerts a force, F, on object B, object B will exert an equal and opposite force on object A. It is important to recognise that the action force and the reaction force in Newton's third law act on different objects and so should never be added together; their effect will only be on the object on which they act. Newton's third law applies not only to forces between objects which are in direct contact, but also to non-contact forces, such as the force due to gravity between objects.

The main misconception that arises when considering Newton's third law is the belief that, if a large mass collides with a smaller mass, then the larger object exerts a larger force and the smaller object exerts a smaller force. This is not true. If you witnessed the collision between the car and the bus in Figure 10.3.3(c) on page 349, you would see the car undergoing a large deceleration while the bus undergoes only a small acceleration.

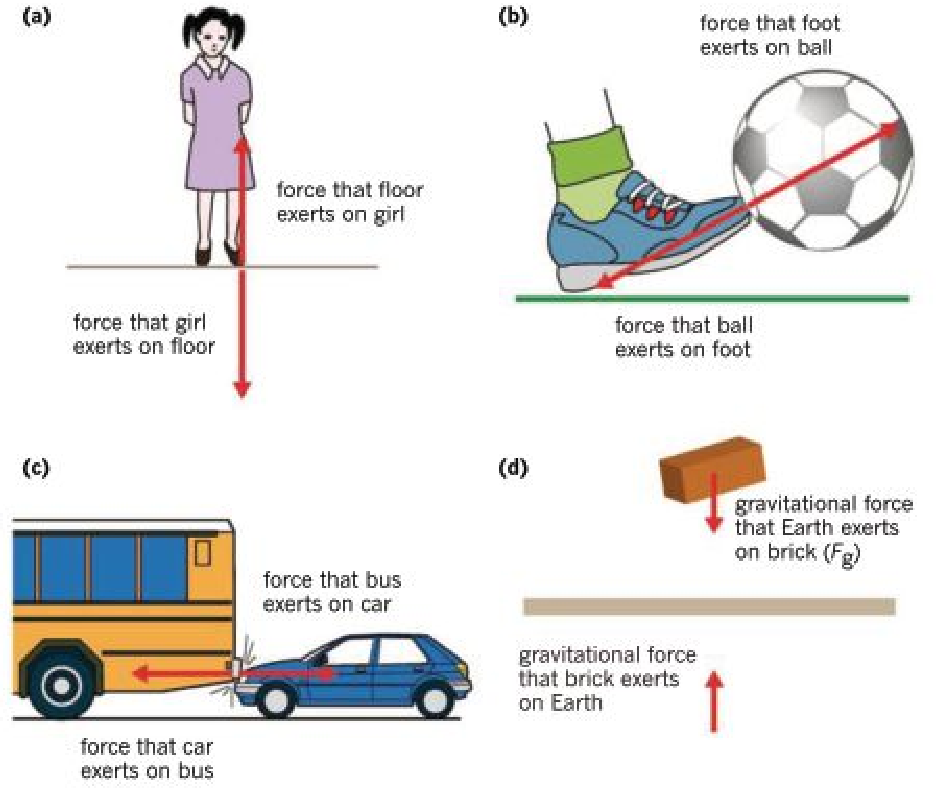


FIGURE 10.3.3 Some action–reaction pairs.

From Newton's second law, you know that the same force acting on a larger mass will result in a smaller acceleration. This is the effect seen in the situation of the car colliding with the bus. Because of the car's small mass, the force acting on the car will cause the car to undergo a large deceleration. The occupants may be seriously injured as a result of this. The force acting on the bus is equal in size, but is acting on a much larger mass. As a result, the bus will have a relatively small acceleration and the occupants will not be as seriously affected.

Identifying the action and reaction forces

When analysing a situation to determine the action and reaction forces according to Newton's third law, it is helpful to be able to label the force vectors systematically. A good strategy for labelling force vectors is to use the capital letter F to represent the force and then to include a subscript consisting of the word 'on' and the thing on which the force is acting, and then the word 'by' and the thing that is applying the force.

The equal and opposite force is then labelled with a capital F and a subscript with the objects in reverse. For example, the action and reaction force vector arrows shown in Figure 10.3.3 can be labelled as shown in Table 10.3.1.

	Action vector	Reaction vector
(a)	$F_{\text{on floor by girl}}$	$F_{\text{on girl by floor}}$
(b)	$F_{\text{on ball by foot}}$	$F_{\text{on foot by ball}}$
(c)	$F_{\text{on bus by car}}$	$F_{\text{on car by bus}}$
(d)	$F_{\text{on Earth by brick}}$	$F_{\text{on brick by Earth}}$

TABLE 10.3.1 Labels of action and reaction force vectors in Figure 10.3.3.

It should be noted that it does not matter which force is considered the action force and which is considered the reaction force. They are always equal in magnitude and opposite in direction.

PHYSICSFILE

Combining Newton's second and third laws in the classroom

You can easily observe the effect of Newton's second and third law in the classroom if you have two dynamics carts with wheels that are free to roll on a smooth surface (such as a bench or desk). If the two carts are placed in contact with each other, and the plunger is activated on *one* of the carts, you will observe that *both* carts roll backwards. This is because of the action and reaction force pair described by Newton's third law. If the two carts have similar masses, you will observe that they accelerate apart at a similar rate. If one cart is heavier than the other, you will observe the lighter cart accelerates at a greater rate. This is because the forces acting on both carts are equal in magnitude and so, according to Newton's second law, the smaller mass will experience a greater acceleration.

Worked example 10.3.1

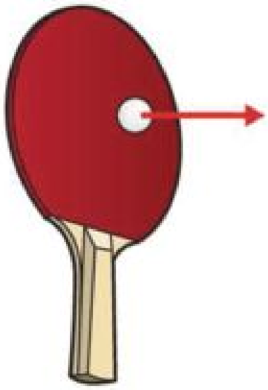
APPLYING NEWTON'S THIRD LAW

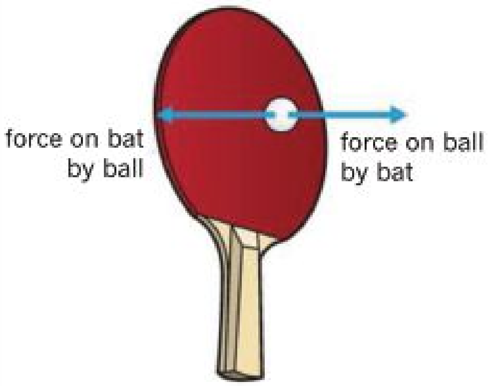
In the diagram below a table-tennis bat is in contact with a table-tennis ball, and one of the forces is given.

a Label the given force using the system 'F_{on} ____________ by ____________'.

b Label the reaction force to the given force using the system 'F_{on} ____________ by ____________'.

c Draw the reaction force on the diagram, showing its size and location.



Thinking	**Working**
Identify the two objects involved in the action–reaction pair.	The bat and the ball.
Identify which object is applying the force and which object is experiencing the force, for the force vector shown.	The force vector shown is a force from the bat on the ball.
Use the system of labelling action and reaction forces 'F_{on} ____________ by ____________' to label the action force.	$F_{\text{on ball by bat}}$
Use the system of labelling action and reaction forces 'F_{on} ____________ by ____________' to label the reaction force.	$F_{\text{on bat by ball}}$
Use a ruler to measure the length of the action force and construct a vector arrow in the opposite direction with its tail on the point of application of the reaction force.	

Worked example: Try yourself 10.3.1

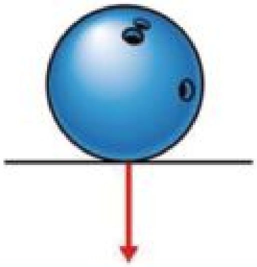
APPLYING NEWTON'S THIRD LAW

In the diagram below, a bowling ball is resting on the floor, and one of the forces is given. Copy the diagram into your book and complete the following:

a Label the given force using the system 'F_{on} __________ by __________'.

b Label the reaction force to the given force using the system 'F_{on} __________ by __________'.

c Draw the reaction force on the diagram, showing its size and location.



NEWTON'S THIRD LAW AND MOTION

Newton's third law also explains how you are able to move around. In fact, Newton's third law is needed to explain all motion. Consider walking. Your leg pushes backwards on the ground with each step. This is an action force on the ground by the foot. As shown in Figure 10.3.4, a component of the force acts downwards and another component pushes backwards horizontally along the surface of the ground.

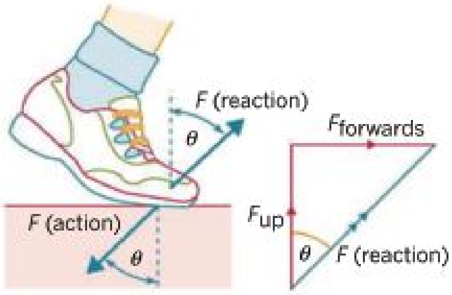


FIGURE 10.3.4 Walking relies on an action and reaction force pair in which the foot pushes down and backwards with an action force. In response, the ground pushes upwards and forwards on you.

The force is transmitted because there is friction between your shoe and the ground. In response, the ground then pushes forwards on you via your foot. This forwards component of the reaction force enables you to move forwards. In other words, it is the ground pushing forwards on you that moves you forwards. It is important to remember that in Newton's second law, $F_{\text{net}} = ma$, the net force, F_{net}, is the sum of the forces acting on the body. This does not include forces that are exerted by the body on other objects. When you push back on the ground, this force is acting on the ground and may affect the ground's motion. If the ground is firm, this effect is usually not noticed, but if you run along a sandy beach, the sand is clearly pushed back by your feet.

The act of walking relies on there being some friction between your shoe and the ground. Without it, there is no grip and it is impossible to supply the action force to the ground. Consequently, the ground cannot supply the reaction force needed to enable forwards motion. Walking on smooth ice is a good example of this. Mountaineers use crampons (basically, a rack of nails) attached to the soles of their boots in order to gain a better grip in icy conditions.

All motion can be explained in terms of action and reaction force pairs. Table 10.3.2 gives some examples of the action and reaction pairs in familiar motions.

Motion	Action force	Reaction force
swimming	hand pushes back on water	water pushes forwards on hand
jumping	legs push down on Earth	Earth pushes up on legs
bicycle or car	tyre pushes back on ground	ground pushes forwards on tyre
jet aircraft and rockets	hot gas is forced backwards out of engine	gases push craft forwards
skydiving	force of gravitation on the skydiver from Earth	force of gravitation on Earth from skydiver

TABLE 10.3.2 Action and reaction force pairs are responsible for all types of motion.

EXTENSION

The inclined plane

In the example of the bin and the table, the table's surface is horizontal. It is possible that an object could be placed on a surface that is tilted so that it makes an angle, *θ*, to the horizontal. In this case, the weight force remains the same: $F_g = mg$ downwards. However, the normal force continues to act at right angles to the surface and will change in magnitude, getting smaller as the angle increases. The magnitude of the normal force is equal in size but opposite in direction to the component of the weight force that acts at right angles to the surface. So, the normal force is $F_N = mg \cos \theta$.

The component of the weight force that acts parallel to the surface will cause the mass to slide down the incline. The motion of the object along the plane will be affected by friction, if it is present. The component of the weight force that acts along the surface is given by $F = mg \sin \theta$.

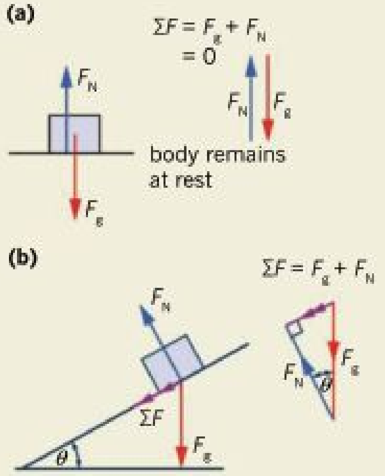


FIGURE 10.3.8 (a) Where the surface is perpendicular to the force due to gravity, the normal force acts directly upwards. (b) On an inclined plane, F_N is at an angle to Fg and is given by $F_N = mg \cos \theta$. If no friction acts, the force that causes the object to accelerate down the plane is $F = mg \sin \theta$.

THE NORMAL FORCE

When an object, for example the rubbish bin shown in Figure 10.3.5, is allowed to fall under the influence of gravity, it is easy to see the effect of the force due to gravity. The action force is the force due to gravity of the Earth on the bin, so the net force on the bin is equal to the force due to gravity, and the bin therefore accelerates at -9.80 m s^{-2}.

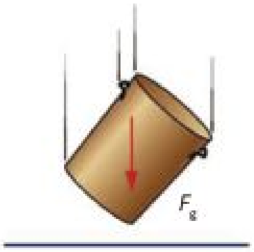


FIGURE 10.3.5 When the bin is in mid-air, there is an unbalanced force due to gravity acting on it so it accelerates towards the ground.

When the bin is at rest on a table, as shown in Figure 10.3.6(a), the force due to gravity ($F_g = mg$) is still acting between the Earth and the bin. Since the bin is at rest, there must be another force acting to balance the force due to gravity on the bin. This upwards force is provided by the table. Because gravity pulls down on the mass of the bin, the bottom of the bin will push down on the surface of the table and the table provides a reaction force on the bin that is equal and opposite, so it will push upwards on the bin as shown in Figure 10.3.6(b).

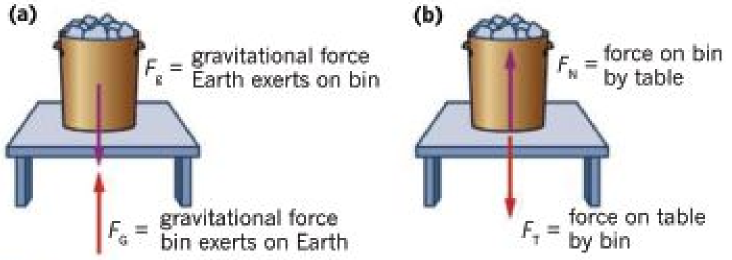


FIGURE 10.3.6 (a) Action–reaction gravitation forces between the bin and the Earth. (b) Action–reaction contact forces between the bin and the table.

The magnitude and direction of the gravitational force on the bin by Earth is equivalent to the magnitude and direction of the force on the table by the bin. Therefore the gravitational force on the bin by Earth is balanced by the upwards contact reaction force on the bin by the table. It is important to note that these two forces are not the pair of forces described in Newton's third law (see Figure 10.3.7). This is because the two forces are both acting on the bin and no pair of Newton's third law force pairs acts on the same object. The contact force provided by a surface that is perpendicular to another surface is called the normal reaction force. It is often abbreviated to normal force, and represented by F_N or N.

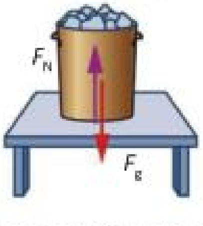


FIGURE 10.3.7 The effect of the two forces on the bin. These are not an action–reaction force pair, even though they are equal in magnitude and opposite in direction.

When you consider only the forces acting on the bin, you are left with the force due to gravity on the bin by Earth and the normal force on the bin by the table. These two forces come from two separate Newton's third law pairs of forces.

10.3 Review

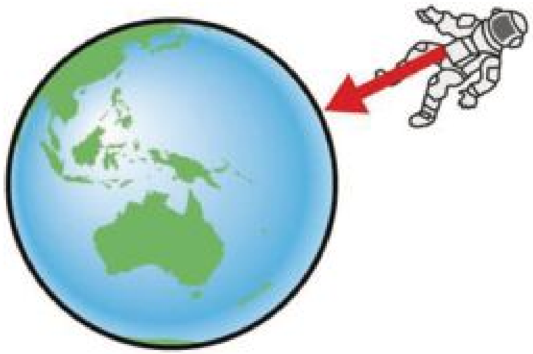
SUMMARY

- For every action (force), there is an equal and opposite reaction (force). This is known as Newton's third law.
- If the action force is labelled systematically, the reaction force can be described by reversing the label of the action force.
- The action and reaction forces are equal and opposite even when the masses of the colliding objects are very different.
- The individual forces making up a Newton's third law pair act on different masses to cause different accelerations according to Newton's second law.
- When an object exerts a downwards force on a surface there is an equal and opposite Newton's third law reaction that exerts a force upwards. This is called the normal force.

KEY QUESTIONS

1 What forces act on a hammer and a nail when a heavy hammer hits a small nail?

2 In the figure below, an astronaut is orbiting the Earth, and one of the forces acting on him is shown by the red arrow.



a Name the given force using the system 'F_{on} ________ by ________'.

b Name the reaction force using the system 'F_{on} ________ by ________'.

3 A swimmer completes a training drill in which he doesn't use his legs to kick, but only uses his stroke to move down the pool. What force causes a swimmer to move forwards down the pool?

4 When an inflated balloon is released it will fly around the room. What is the force that causes the balloon to move?

5 Determine the reaction force involved when a ball is hit with a racquet with a force of 100 N west.

6 A 70 kg fisherman is fishing in a 40 kg dinghy at rest on a still lake when, suddenly, he is attacked by a swarm of wasps. To escape, he leaps into the water and exerts a horizontal force of 140 N north on the boat.

a What force does the boat exert on the fisherman?

b With what acceleration will the boat move initially?

c If the force on the fisherman lasted for 0.50 s, determine the initial speed attained by both the man and the boat.

7 An astronaut becomes untethered during a space-walk and drifts away from the spacecraft. To get back to the ship, she decides to throw her tool kit away. In which direction should she throw the tool kit?

8 Two students, James and Tania, are discussing the forces acting on a lunchbox that is sitting on the laboratory bench. James states that the weight force and the normal force are acting on the lunchbox and that since these forces are equal in magnitude but opposite in direction, they comprise a Newton's third law action–reaction pair. Tania disagrees, saying that these forces are not an action–reaction pair. Who is correct and why?

10.4 Momentum and conservation of momentum

It is possible to understand some physics concepts intuitively without knowing the physics terms or words that describe them. For example, you may know that once a heavy object gets moving it is difficult to stop it, whereas a lighter object moving at the same speed is easier to stop. In the previous sections of this chapter you have seen how Newton's laws of motion can be used to explain these observations. In this section you will explore how these observations can be related to the concept called momentum.

MOMENTUM

The **momentum** of an object relates to both its mass and its velocity. The footballers colliding in Figure 10.4.1 have momentum due to their mass, and the faster they run, the more momentum they will have. A slower moving player has less momentum than one who is moving faster. Similarly, a person with greater mass will have more momentum than a smaller, lighter person travelling at the same speed. The more momentum an object has, due to its mass or its velocity, the more momentum it has to lose before it stops.

FIGURE 10.4.1 Momentum is related to mass and velocity. The greater the mass or velocity, the harder it is to stop or start moving.

The equation for momentum, p, is the product of the object's mass, m, and its velocity, v.

$p = mv$

where p is momentum (kg m s^{-1})

m is the mass of the object (kg)

v is the velocity of the object (m s^{-1}).

The greater an object's mass or velocity, the larger that object's momentum will be. As velocity is a vector quantity, momentum is also a vector and so it must have magnitude, units and direction. The direction of a momentum vector will always be the same as the direction of the velocity vector. For calculations of change in momentum in a single dimension, we can use the sign conventions of positive and negative.

Force is equal to the rate of change of momentum. This can be mathematically explained using Newton's second law, which was presented in Section 10.2 on page 340.

The following derivation will show how Newton's second law relates to momentum. It results with net force, F_{net}, equal to the change in momentum, Δp, divided by the period of time, Δt, which is the rate of change of momentum:

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

$$F_{net} = ma$$

$$F_{net} = m\frac{(v-u)}{\Delta t}$$

$$F_{net} = \frac{mv - mu}{\Delta t}$$

$$F_{net} = \frac{\Delta p}{\Delta t}$$

This means that changes in momentum are caused by the action of a net force.

Worked example 10.4.1

MOMENTUM

Calculate the momentum of a 60.0 kg student walking at 3.5 m s^{-1} east.

Thinking	Working
Ensure that the variables are in their standard units.	$m = 60.0$ kg $v = 3.50$ m s^{-1} east
Apply the equation for momentum.	$p = mv$ $= 60.0 \times 3.50$ $= 210$ kg m s^{-1} east

Worked example: Try yourself 10.4.1

MOMENTUM

Calculate the momentum of a 1230 kg car driving at 16.7 m s^{-1} north.

CONSERVATION OF MOMENTUM

The most significant feature of momentum is that it is **conserved** in any interaction or collision between objects. This means that the total (sum of) momentum in any system before a collision will be equal to the total (sum of) momentum in the system after the collision. This is known as the law of conservation of momentum and can be represented by the following relationship:

$$\sum p_{before} = \sum p_{after}$$

where $\sum p$ is the sum of the momentum of objects in a system.

To find the total momentum of objects in a system (either before or after a collision) simply find the momentum of each object, considering their masses and velocities, and then add them together.

For collisions in one dimension, apply the sign convention of positive and negative directions to the velocities and then use algebra to determine the answer to the problem. For collisions in two dimensions, resolve vectors describing motion into perpendicular components and then consider the conservation of momentum in each single dimension.

PHYSICSFILE

The discovery of the neutron

The law of conservation of momentum was used to interpret the data from investigations that led to the discovery of the neutron. Because the neutron has no charge, it could not be investigated through the interactions of charged particles that had led to the discovery of the proton and electron. In 1932, James Chadwick investigated collisions between alpha particles and the element beryllium. However, the conservation of momentum calculations didn't add up. Chadwick knew that the law of conservation of momentum was true, so he reasoned that there was an unknown particle involved that had a mass close to the proton's mass, but without electric charge. Subsequent investigations confirmed his experiments and led to the naming of this particle as the neutron.

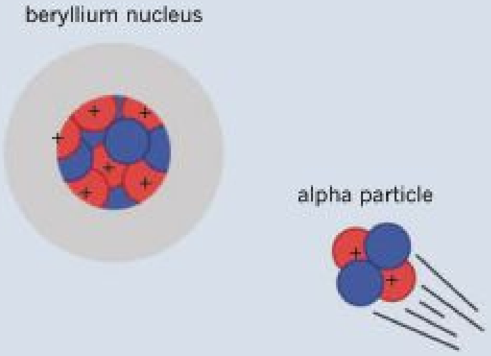


FIGURE 10.4.2 Investigating collisions between alpha particles and a beryllium nucleus led to the discovery of the neutron.

Momentum in one-dimensional collisions

If two objects are colliding in one dimension, then the following equation applies:

$$\sum p_{\text{before}} = \sum p_{\text{after}}$$

$$m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2$$

where m_1 is the mass of object 1 (kg)

u_1 is the initial velocity of object 1 (m s^{-1})

v_1 is the final velocity of object 1 (m s^{-1})

m_2 is the mass of object 2 (kg)

u_2 is the initial velocity of object 2 (m s^{-1})

v_2 is the final velocity of object 2 (m s^{-1}).

PHYSICS IN ACTION

Elastic and inelastic collisions

Collisions can either be elastic or inelastic. In elastic collisions no kinetic energy is lost, while inelastic collisions will result in some kinetic energy being converted to another form, such as thermal energy. The loss in kinetic energy during inelastic collisions will result in a lower combination of final velocities. If the collision was 100% inelastic, then all of the kinetic energy would be converted to thermal energy and both objects would be stationary immediately after the collision. If the collision was 100% elastic, then the final velocities would be maximum.

It is impossible to predict the final velocities of two objects colliding unless the collision is perfectly elastic, as it is not possible to predict the kinetic energy lost in any collision.

Worked example 10.4.2

CONSERVATION OF MOMENTUM

A 2.50 kg mass is moving at 4.50 m s^{-1} west towards a 1.50 kg mass moving at 3.00 m s^{-1} east. Calculate the final velocity of the 2.50 kg mass if the 1.50 kg mass rebounds at 5.00 m s^{-1} west.

Thinking	Working
Identify the variables using subscripts. Ensure that the variables are in their standard units.	$m_1 = 2.50$ kg $u_1 = 4.50$ m s^{-1} west $v_1 = ?$ $m_2 = 1.50$ kg $u_2 = 3.00$ m s^{-1} east $v_2 = 5.00$ m s^{-1} west
Apply the sign convention to the variables.	$m_1 = 2.50$ kg $u_1 = -4.50$ m s^{-1} $v_1 = ?$ $m_2 = 1.50$ kg $u_2 = +3.00$ m s^{-1} $v_2 = -5.00$ m s^{-1}
Apply the equation for conservation of momentum.	$\sum p_{\text{before}} = \sum p_{\text{after}}$ $m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2$ $(2.50 \times -4.50) + (1.50 \times 3.00)$ $= 2.50v_1 + (1.50 \times -5.00)$ $2.50v_1 = -11.25 + 4.50 - (-7.50)$ $v_1 = \frac{0.75}{2.50}$ $= 0.30$ m s^{-1}
Apply the sign convention to describe the direction of the final velocity.	$v_1 = 0.30$ m s^{-1} east

Worked example: Try yourself 10.4.2

CONSERVATION OF MOMENTUM

A 1200 kg wrecking ball is moving at 2.50 m s^{-1} north towards a 1500 kg wrecking ball moving at 4.00 m s^{-1} south. Calculate the final velocity of the 1500 kg ball if the 1200 kg ball rebounds at 3.50 m s^{-1} south.

PHYSICS IN ACTION

Conservation of momentum in sports

The law of conservation of momentum has applications in many spheres of human endeavour. In sport, however, the law is particularly evident. Consider the following sports:

- curling
- lawn bowls
- tenpin bowling
- bocce
- pool
- snooker
- marbles.

FIGURE 10.4.3 Curling is a sport intrinsically linked to the conservation of momentum.

All of these sports require the athlete to cause one object to collide with the other. The best players are able to control the initial velocity of the striking object such that the magnitude and direction of momentum for both the striking object and the target object after the collision result in a good score or a positional advantage for themselves or their team.

FIGURE 10.4.4 Controlling the direction and momentum of this bowling ball will enable to player to pick up a spare.

Momentum when masses combine

It is important to note that in the situations described in Worked example 10.4.2 (p. 357), the two objects remain separate from each other. However, it is possible for two objects to combine (stick together) when they collide. If two objects combine when they collide, then the equation is modified to:

$$\sum p_{\text{before}} = \sum p_{\text{after}}$$

$$m_1u_1 + m_2u_2 = m_3v_3$$

where m_1 is the mass of object 1 (kg)

u_1 is the initial velocity of object 1 (m s^{-1})

m_2 is the mass of object 2 (kg)

u_2 is the initial velocity of object 2 (m s^{-1})

m_3 is the combined mass of m_1 and m_2 (kg)

v_3 is the final velocity of combined mass of m_1 and m_2 (m s^{-1}).

Worked example 10.4.3

CONSERVATION OF MOMENTUM WHEN MASSES COMBINE

A 5.00 kg lump of clay is moving at 2.00 m s^{-1} west towards a 7.50 kg mass of clay moving at 3.00 m s^{-1} east. They collide to form a single, combined mass of clay. Calculate the final velocity of the combined mass of clay.

Thinking	Working
Identify the variables using subscripts and ensure that the variables are in their standard units. Add m_1 and m_2 to get m_3.	$m_1 = 5.00$ kg $u_1 = 2.00$ m s^{-1} west $m_2 = 7.50$ kg $u_2 = 3.00$ m s^{-1} east $m_3 = 12.50$ kg $v_3 = ?$
Apply the sign convention to the variables.	$m_1 = 5.00$ kg $u_1 = -2.00$ m s^{-1} $m_2 = 7.50$ kg $u_2 = +3.00$ m s^{-1} $m_3 = 12.50$ kg $v_3 = ?$
Apply the equation for conservation of momentum.	$\sum p_{\text{before}} = \sum p_{\text{after}}$ $m_1u_1 + m_2u_2 = m_3v_3$ $(5.00 \times -2.00) + (7.50 \times 3.00) = 12.50v_3$ $v_3 = \frac{-10.0 + 22.50}{12.50}$ $= 1.00$ m s^{-1}
Apply the sign convention to describe the direction of the final velocity.	$v_3 = 1.00$ m s^{-1} east

Worked example: Try yourself 10.4.3

CONSERVATION OF MOMENTUM WHEN MASSES COMBINE

An 80.0 kg rugby player is moving at 1.50 m s^{-1} north when he tackles an opponent with a mass of 50.0 kg who is moving at 5.00 m s^{-1} south. Calculate the final velocity of the two players.

Momentum in explosive collisions

It is also possible for one object to break apart into two objects in what is known as an 'explosive collision'. If one object breaks apart when an explosive collision occurs, then the equation is modified to:

$$\sum p_{\text{before}} = \sum p_{\text{after}}$$

$$m_1u_1 = m_2v_2 + m_3v_3$$

where m_1 is the mass of object 1 (2 and 3 combined) (kg)

u_1 is the initial velocity of object 1 (m s^{-1})

m_2 is the mass of object 2 (kg)

v_2 is the final velocity of object 2 (m s^{-1})

m_3 is the mass of object 3 (kg)

v_3 is the final velocity of object 3 (m s^{-1}).

Worked example 10.4.4

CONSERVATION OF MOMENTUM FOR EXPLOSIVE COLLISIONS

A 90.0 kg athlete holds a 1000 g javelin. She approaches the line at 7.75 m s^{-1} west and releases the javelin down the field. After throwing it, she continues with a velocity of 7.25 m s^{-1} west. Calculate the velocity of the javelin just after she releases it.

Thinking	Working
Identify the variables using subscripts and ensure that the variables are in their standard units. Note that m_1 is the sum of the bodies, i.e. the athlete and the javelin.	$m_1 = 91$ kg $u_1 = 7.75$ m s^{-1} west $m_2 = 90$ kg $v_2 = 7.25$ m s^{-1} west $m_3 = 1.00$ kg $v_3 = ?$
Apply the sign convention to the variables.	$m_1 = 91$ kg $u_1 = -7.75$ m s^{-1} $m_2 = 90$ kg $v_2 = -7.25$ m s^{-1} $m_3 = 1.00$ kg $v_3 = ?$
Apply the equation for conservation of momentum for explosive collisions.	$\sum p_{\text{before}} = \sum p_{\text{after}}$ $m_1u_1 = m_2u_2 + m_3v_3$ $91.0 \times -7.75 = (90.0 \times -7.25) + 1.00v_3$ $v_3 = \frac{-705.25 - (-652.5)}{1.00}$ $v_3 = \frac{-52.75}{1.00}$ $= -52.8$ m s^{-1}
Apply the sign convention to describe the direction of the final velocity.	$v_3 = 52.8$ m s^{-1} west

Worked example: Try yourself 10.4.4

CONSERVATION OF MOMENTUM FOR EXPLOSIVE COLLISIONS

A 2000 kg cannon fires a 10 kg cannonball. The cannon and the cannonball are initially stationary. After firing, the cannon recoils with a velocity of 8.15 m s^{-1} north. Calculate the velocity of the cannonball just after it is fired.

PHYSICSFILE

Conservation of momentum in engines

If you release an inflated rubber balloon with its neck open, it will fly off around the room. In the diagram below, the momentum of the air to the left results in the movement of the balloon to the right. Momentum is conserved.

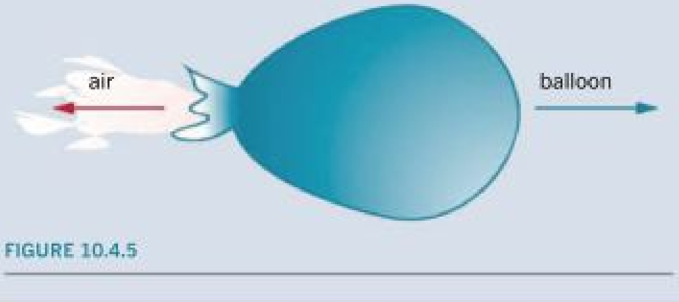


FIGURE 10.4.5

This is the principle upon which rockets and jet engines are based. Both rockets and jet engines employ a high-velocity stream of hot gases that are vented after the combustion of a fuel–air mixture. The hot exhaust gases have a very large momentum as a result of the high velocities involved, and can accelerate rockets and jets to high velocities as they acquire an equal momentum in the opposite direction. Rockets destined for space carry their own oxygen supply, while jet engines use the surrounding air supply.

10.4 Review

SUMMARY

- Momentum is the product of an object's mass and velocity.
- Momentum is a vector quantity and is calculated using the equation: $p = mv$
- Force is equal to the rate of change of momentum.
- The law of conservation of momentum can be applied to situations in which:
  - two objects collide and remain separate:
    $\sum p_{\text{before}} = \sum p_{\text{after}}$
    $m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2$
  - two objects collide and combine together:
    $\sum p_{\text{before}} = \sum p_{\text{after}}$
    $m_1u_1 + m_2u_2 = m_3v_3$
  - one object breaks apart into two objects in an explosive collision:
    $\sum p_{\text{before}} = \sum p_{\text{after}}$
    $m_1u_1 = m_2u_2 + m_3v_3$

KEY QUESTIONS

1 Calculate the momentum of a 3.50 kg fish swimming at 2.50 m s^{-1} south.

2 Calculate the momentum of a 433 kg boat travelling at 22.2 m s^{-1} west.

3 Calculate the momentum of a 65.0 g tennis ball being served at 61.0 m s^{-1} south.

4 Which object has the greater momentum: a medicine ball of mass 4.5 kg travelling at 3.5 m s^{-1} or one of mass 2.5 kg travelling at 6.8 m s^{-1}?

5 A rower of mass 70.0 kg steps out of a stationary boat with a velocity of 2.50 m s^{-1} forwards onto the nearby riverbank. The boat has a mass of 400 kg and was initially at rest. With what velocity does the boat begin to move as the rower steps out? Give the answer to three significant figures.

6 A golf ball of mass 70.0 g is stationary on the ground when it is hit by a 545 g golf club travelling at 80.0 m s^{-1}. If the ball leaves the club at a speed of 75.0 m s^{-1}, with what speed does the club move just after hitting the ball? Give the answer to three significant figures.

7 A railway wagon of mass 2.50 tonnes moving along a horizontal track at 2.00 m s^{-1} runs into a stationary engine and is coupled to it. After the collision, the engine and wagon move off at a slow 0.300 m s^{-1}. What is the mass of the engine alone? Give the answer to three significant figures.

8 A space shuttle of mass 10 000 kg, initially at rest, burns 5.0 kg of fuel and oxygen in its rockets to produce exhaust gases ejected at a velocity of 6000 m s^{-1}. Calculate the velocity that this exchange will give to the space shuttle.

9 A small research rocket of mass 250 kg is launched vertically as part of a weather study. It sends out 50 kg of burnt fuel and exhaust gases with a velocity of 180 m s^{-1} in a 2 s initial acceleration period.

  a What is the velocity of the rocket after this initial acceleration?

  b What upwards force does this apply to the rocket?

  c What is the net upwards acceleration acting on the rocket? (Use $g = 10$ m s^{-2} if required.)

10.5 Momentum transfer

The previous section described the momentum of an object in terms of its velocity and its mass. For each of the different collisions described in that section, the momentum of the system was conserved. That is, when all of the objects involved in the collision were considered, the total momentum before and after the collision was the same. But for each separate object, considered in isolation, momentum may not have been conserved. In the examples explored, an object experienced a change in its velocity due to the collision.

When an object changes its velocity, its momentum will also change. An increase in velocity means an increase in momentum, while a decrease in velocity corresponds to a decrease in momentum. Change in momentum, Δp, is also called **impulse**, I.

You can also consider this change in momentum as a transfer of momentum. For this reason, momentum transfer is often referred to as impulse, I.

CHANGE IN MOMENTUM IN ONE DIMENSION

It is easy to change the velocity of an object. You can either run faster or run slower; you can press a little harder on the pedals of a bike or press a little softer. You can also bounce an object off a surface. For example, the basketball in Figure 10.5.1 experiences a change in velocity when it changes direction during the bounce. The cause of these changes in motion will be discussed in the Section 10.6. First, consider impulse or change in momentum in one dimension.

FIGURE 10.5.1 A bouncing basketball undergoes a change in momentum when it changes direction as it bounces.

The term 'impulse' means change in momentum. So the impulse or change in momentum of an object moving in one dimension is calculated using the equation:

$$\begin{aligned} I &= \Delta p \\ &= p_{\text{final}} - p_{\text{initial}} \\ &= mv - mu \end{aligned}$$

where I is the impulse (kg m s^{-1})

Δp is the change in momentum (kg m s^{-1})

m is the mass (kg)

v is the final velocity (m s^{-1})

u is the initial velocity (m s^{-1}).

As momentum is a vector quantity, the impulse or change in momentum is also a vector, so it is expressed in magnitude, units and direction.

Worked example 10.5.1

IMPULSE OR CHANGE IN MOMENTUM

A student rides a bike to school and approaches the bike rack at 8.20 m s^{-1} east. Calculate the impulse of the student during the time it takes to stop if the student and the bike have a combined mass of 80 kg and the student stops at the rack.

Thinking	Working
Ensure that the variables are in their standard units.	$m = 80$ kg $u = 8.20$ m s^{-1} east $v = 0$ m s^{-1}
Apply the sign convention to the velocity vector.	$m = 80$ kg $u = +8.20$ m s^{-1} $v = 0$ m s^{-1}
Apply the equation for impulse or change in momentum.	$I = mv - mu$ $= (80 \times 0) - (80 \times 8.20)$ $= 0 - 656$ $= -656$ kg m s^{-1}
Apply the sign convention to describe the direction of the impulse.	$I = 656$ kg m s^{-1} west

Worked example: Try yourself 10.5.1

IMPULSE OR CHANGE IN MOMENTUM

A student hurries to class after lunch, moving at 4.55 m s^{-1} north. Suddenly the student remembers that she has forgotten her laptop and goes back to her locker at 6.15 m s^{-1} south. If her mass is 75.0 kg, calculate the impulse of the student during the time it takes to turn around.

CHANGE IN MOMENTUM IN TWO DIMENSIONS

The velocity of an object can be changed not only by changing the magnitude of its velocity, but also by changing the direction of its motion. The velocity of the boat in Figure 10.5.2, for example, changes because the boat changes direction. As you saw in Chapter 8 'Scalars and vectors', a change in velocity in two dimensions can be calculated using geometry. The equation for impulse can be manipulated slightly to illustrate where the change in velocity is applied:

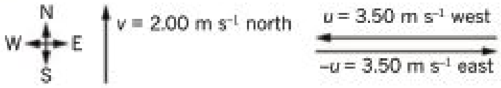
$$\begin{aligned} I &= mv - mu \\ &= m(v - u) \end{aligned}$$

FIGURE 10.5.2 Changing momentum in two dimensions by changing direction.

Worked example 10.5.2

IMPULSE OR CHANGE IN MOMENTUM IN TWO DIMENSIONS

A 65.0 kg mass is moving at 3.50 m s^{-1} west and then changes to 2.00 m s^{-1} north. Calculate the change in momentum of the mass over the period of the change.

Thinking	**Working**
Identify the formula for calculating a change in velocity, Δv.	Δv = final velocity – initial velocity
Draw the final velocity vector, *v*, and the initial velocity vector, *u*, separately. Then draw the initial velocity in the opposite direction, which represents the negative of the initial velocity, –*u*.	
Construct a vector diagram drawing *v* first and then from its head draw the opposite of *u*. The change of velocity vector is drawn from the tail of the final velocity to the head of the opposite of the initial velocity.	
As the two vectors to be added are at 90° to each other, apply Pythagoras' theorem to calculate the magnitude of the change in velocity.	$\Delta v^2 = 2.00^2 + 3.50^2$ $= 4.00 + 12.25$ $\Delta v = \sqrt{16.25}$ $= 4.03 \text{ m s}^{-1}$
Calculate the angle from the north vector to the change in velocity vector.	$\tan\theta = \frac{3.50}{2.00}$ $\theta = \tan^{-1}1.75$ $= 60.3°$
State the magnitude and direction of the change in velocity.	$\Delta v = 4.03 \text{ m s}^{-1}$ N 60.3° E
Identify the variables using subscripts and ensure that the variables are in their standard units.	$m = 65.0$ kg $\Delta v = 4.03 \text{ m s}^{-1}$ N 60.3° E
Apply the equation for impulse or change in momentum.	$\Delta p = mv - mu$ $= m(v - u)$ $= m\Delta v$ $= 65.0 \times 4.03$ $= 262 \text{ kg m s}^{-1}$
Apply the direction convention to describe the direction of the change in momentum.	$\Delta p = 262 \text{ kg m s}^{-1}$ N 60.3° E

Worked example: Try yourself 10.5.2

IMPULSE OR CHANGE IN MOMENTUM IN TWO DIMENSIONS

A 65.0 g pool ball is moving at 0.250 m s^{-1} south towards a cushion and bounces off at 0.200 m s^{-1} east. Calculate the impulse on the ball during the change in velocity.

10.5 Review

SUMMARY

- Change or transfer in momentum, Δp, is also known as impulse, I. It is a vector quantity.
- A change or transfer in momentum occurs when an object changes its velocity.
- The equation for impulse is: $I = \Delta p = mv - mu$
- Change in momentum in two directions can be calculated using geometry.

KEY QUESTIONS

1 Calculate the impulse of a 9.50 kg dog that changes its velocity from 2.50 m s^{-1} north to 6.25 m s^{-1} south.

2 Calculate the impulse of a 6050 kg truck as it changes from moving at 22.2 m s^{-1} west to 16.7 m s^{-1} east.

3 The velocity of an 8.00 kg mass changes from 3.00 m s^{-1} east to 8.00 m s^{-1} east. Calculate the change in momentum.

4 Calculate the change in momentum of a 250 g apple as it changes from rest to moving downwards at 9.80 m s^{-1} after falling off a tree.

5 The momentum of a ball of mass 0.125 kg changes by 0.075 kg m s^{-1} south. If its original velocity was 3.00 m s^{-1} north, what is the final velocity?

6 A 45.0 kg mass moving at 45.0 m s^{-1} west changes direction so that it moves at 45.0 m s^{-1} north. Calculate the change in momentum of the mass over the period of the change.

7 A marathon runner with a mass of 70.0 kg is running with a velocity of 4.00 m s^{-1} north, and then turns a corner to start running 3.60 m s^{-1} west. Calculate the marathon runner's change in momentum.

10.6 Momentum and net force

Section 10.2 on Newton's second law of motion discussed the quantitative connection between force, mass, time and change in velocity. This relationship is explored further in this section. The relationship between change in momentum, Δp, the period of time, Δt, and net force, F_{net}, helps to explain the effects of collisions and how to minimise those effects. It is the key to providing safer environments, including in sporting contexts such as that shown in Figure 10.6.1.

FIGURE 10.6.1 When two footballers collide, they exert an equal and opposite force on each other.

Think about what it would feel like to fall onto a concrete floor. Even from a small height it would hurt. A fall from the same height onto a foam mattress would barely be felt. In both situations speed is the same, mass has not changed and gravity provides the same acceleration, no matter the mass. Yet each experience would feel different.

CHANGE IN MOMENTUM (IMPULSE)

According to Newton's second law, a net force will cause a mass to accelerate. A larger net force will create a faster change in the velocity of the mass. The faster the change occurs – that is, the smaller the period of time, Δt – the greater the net force that produced that change. Landing on a concrete floor changes the velocity of an object very quickly. The falling object is brought to an abrupt stop within a very short amount of time. When landing on a foam mattress, the change occurs over a much greater timeframe. Therefore, the force needed to produce the change is smaller.

Starting with the equation introduced in Section 10.4, the relationship between change in momentum, Δp, or impulse, I, and the variables of force, F_{net} (often written just as F), and period of time, Δt, becomes:

$$F_{net} = \frac{\Delta p}{\Delta t}$$

$$= \frac{mv - mu}{\Delta t}$$

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

$$F_{net}\Delta t = m(v - u)$$

$$= I$$

where I is the impulse (kg m s^{-1}).

These equations illustrate that for a given change in momentum or impulse, the product of force and period of time is constant. This relationship is key to understanding collisions. Worked examples 10.6.1 and 10.6.2, below, illustrate how this works.

Worked example 10.6.1

CALCULATING THE FORCE AND IMPULSE

A student drops a 105 g pool ball onto a concrete floor from a height of 2.00 m. Just before it hits the floor, the velocity of the ball is 6.26 m s^{-1} down. Before it bounces back up, there is an instant in time at which the ball's velocity is zero. The time it takes for the ball to change its velocity to zero is 5.02 milliseconds.

a Calculate the change in momentum of the pool ball.	
Thinking	**Working**
Ensure that the variables are in their standard units.	$m = 0.105$ kg $u = 6.26$ m s^{-1} down $v = 0$ m s^{-1}
Apply the sign and direction convention for motion in one dimension. Up is positive and down is negative.	$m = 0.105$ kg $u = -6.26$ m s^{-1} $v = 0$ m s^{-1}
Apply the equation for change in momentum.	$\Delta p = m(v - u)$ $= 0.105 \times (0 - (-6.26))$ $= 0.657$ kg m s^{-1}
Refer to the sign and direction convention to determine the direction of the change in momentum.	$\Delta p = 0.657$ kg m s^{-1} up

b Calculate the impulse of the pool ball.	
Thinking	**Working**
Using the answer to part (a), apply the equation for impulse.	$I = \Delta p$ $= 0.657$ kg m s^{-1}
Refer to the sign and direction convention to determine the direction of the impulse.	$I = 0.657$ kg m s^{-1} up

c Calculate the average force that acts to cause the impulse.	
Thinking	**Working**
Use the answer to part (b). Ensure that the variables are in their standard units.	$I = 0.657$ kg m s^{-1} $\Delta t = 5.02 \times 10^{-3}$ s
Apply the equation for force.	$F\Delta t = I$ $F = \frac{I}{\Delta t}$ $= \frac{0.657}{5.02 \times 10^{-3}}$ $= 131$ N
Refer to the sign and direction convention to determine the direction of the force.	$F = 131$ N up

Worked example: Try yourself 10.6.1

CALCULATING THE FORCE AND IMPULSE

A student drops a 56.0 g egg onto a table from a height of 60 cm. Just before it hits the table, the velocity of the egg is 3.43 m s^{-1} down. The egg's final velocity is zero as it smashes on the table. The time it takes for the egg to change its velocity to zero is 3.55 milliseconds.

a Calculate the change in momentum of the egg.

b Calculate the impulse of the egg.

c Calculate the average force that acts to cause the impulse.

Worked example 10.6.2

CALCULATING THE FORCE AND IMPULSE (SOFT LANDING)

A student drops a 105 g pool ball onto a foam mattress from a height of 2.00 m. Just before it hits the foam mattress, the velocity of the ball is 6.26 m s^{-1} down. Before it bounces back up, there is an instant in time at which the ball's velocity is zero. The time it takes for the ball to change its velocity to zero is 0.360 seconds.

a Calculate the change in momentum of the pool ball.	
Thinking	**Working**
Ensure that the variables are in their standard units.	$m = 0.105$ kg $u = 6.26$ m s^{-1} down $v = 0$ m s^{-1}
Apply the sign and direction convention for motion in one dimension. Up is positive and down is negative.	$m = 0.105$ kg $u = -6.26$ m s^{-1} $v = 0$ m s^{-1}
Apply the equation for change in momentum.	$\Delta p = m(v - u)$ $= 0.105 \times (0 - (-6.26))$ $= 0.657$ kg m s^{-1}
Refer to the sign and direction convention to determine the direction of the change in momentum.	$\Delta p = 0.657$ kg m s^{-1} up

b Calculate the impulse of the pool ball.

Thinking	Working
Using the answer to part (a), apply the equation for impulse.	$I = \Delta p$ $= 0.657\ \text{kg m s}^{-1}$
Refer to the sign and direction convention to determine the direction of the impulse.	$I = 0.657\ \text{kg m s}^{-1}$ up

c Calculate the average force that acts to cause the impulse.

Thinking	Working
Using the answer to part (b), ensure that the variables are in their standard units.	$I = 0.657\ \text{kg m s}^{-1}$ $\Delta t = 0.360\ \text{s}$
Apply the equation for force.	$F\Delta t = I$ $F = \frac{I}{\Delta t}$ $= \frac{0.657}{0.360}$ $= 1.83\ \text{N}$
Refer to the sign and direction convention to determine the direction of the force.	$F = 1.83$ N up

Worked example: Try yourself 10.6.2

CALCULATING THE FORCE AND IMPULSE (SOFT LANDING)

A student drops a 56.0 g egg into a mound of flour from a height of 60 cm. Just before it hits the mound of flour, the velocity of the egg is 3.43 m s^{-1} down. The egg's final velocity is zero as it sinks into the mound of flour. The time it takes for the egg to change its velocity to zero is 0.325 seconds.

a Calculate the change in momentum of the egg.

b Calculate the impulse of the egg.

c Calculate the average force that acts to cause the impulse.

From these worked examples you should notice a number of important things:

- The change in momentum and the impulse were always the same.
- Regardless of the surface that the object landed on, the impulse or change in momentum remained the same.
- The period of time was the main difference between the two different surfaces. Hard surfaces resulted in a short time to stop and soft surfaces resulted in a longer time to stop.
- The effect of the period of time on the force was dramatic. A shorter time meant a greater force, while a longer time meant a much smaller force.

DETERMINING IMPULSE FROM A CHANGING FORCE

In the previous examples it was assumed that the force that acted to change the impulse over a period of time was constant during that time. This is not always the case in real situations. Often the force varies over the period of the impact, so there needs to be a way to determine the impulse as the force varies.

An illustration of this is when a tennis player strikes a ball with a racquet. At the instant the ball comes in contact with the racquet, the applied force will be small. As the strings distort and the ball compresses, the force will increase until the ball has been stopped. The force will then decrease as the ball accelerates away from the racquet. A graph of force against time is shown in Figure 10.6.2.

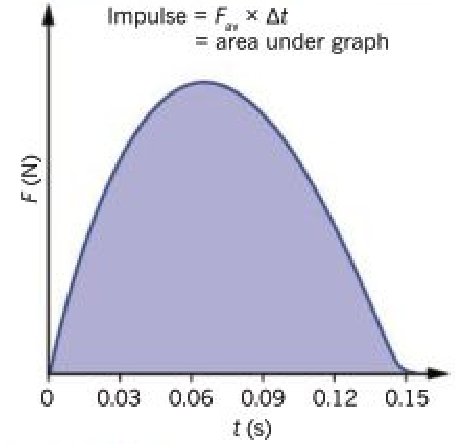


FIGURE 10.6.2 The forces acting on the tennis ball during its collision with the racquet are not constant.

The impulse, I, affecting the ball during any time interval will be the product of applied force, F, and the period of time, Δt. The total impulse during the period of time the ball is in contact with the racquet will be:

$$I = F_{av}\Delta t$$

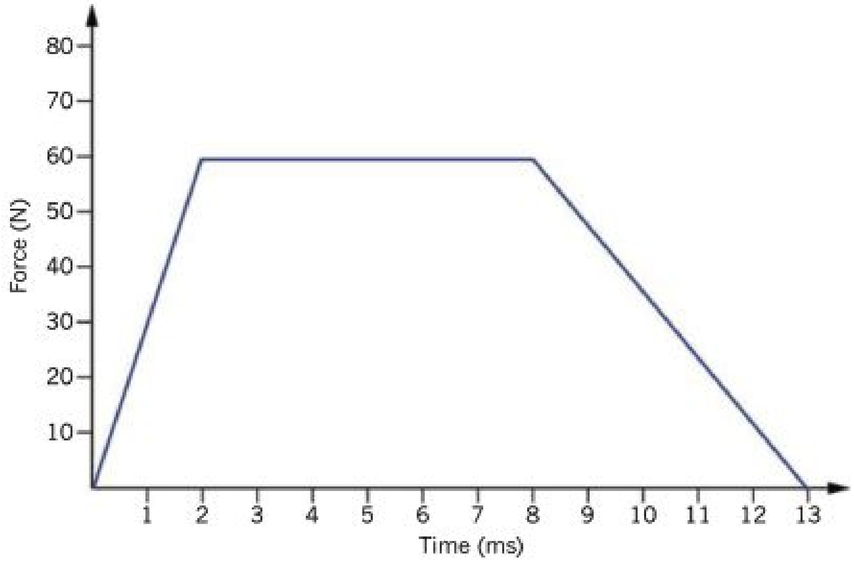
where F_{av} is the average force applied during the collision and Δt is the total period of time the ball is in contact with the racquet. In a graph showing force against time, the area under the line is a product of the height (force) and the width (time). Thus, the total area under the line in a force against time graph is the total impulse for any collision, even those in which the force is not constant.

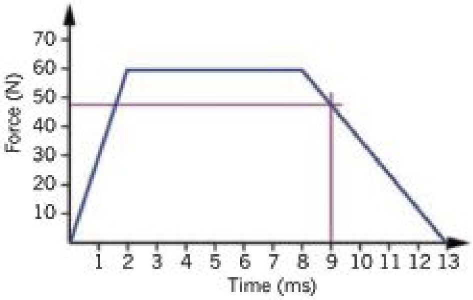
The concept of impulse is appropriate when dealing with forces during any collision since it links force and contact time as, for example, when a person's foot hits the ground or when a ball is hit by a bat or racquet. If applied to situations where contact is over an extended period of time, the average net force involved is used since the forces are generally changing (as the ball deforms, for example). The average net applied force can be found directly from the formula for impulse. The instantaneous applied force at any particular time during the collision must be read from a graph of force against time.

Worked example 10.6.3

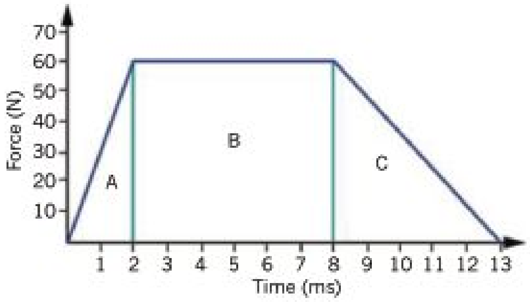
CALCULATING THE TOTAL IMPULSE FROM A CHANGING FORCE

A student records the force acting on a rubber ball as it bounces off a hard concrete floor over a period of time. The graph shows the forces acting on the ball during its collision with the concrete floor.



a Determine the force acting on the ball at a time of 9.0 milliseconds.	
Thinking	**Working**
From the 9.0 millisecond point on the x-axis go up to the line of the graph, then across to the y-axis.	
The force is estimated by reading the intercept of the y-axis.	$F = 48$ N

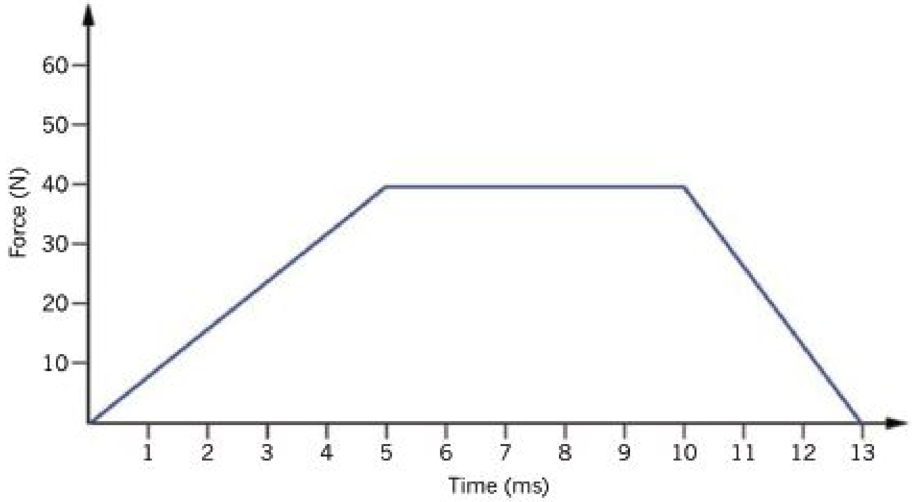
b Calculate the total impulse of the ball over the 13 millisecond period of time.

Thinking	Working
Break the area under the graph into sections for which you can calculate the area.	In this case, the graph can be broken into three sections: A, B and C. 
Calculate the area of the three sections *A*, *B* and *C* using the equations for area of a triangle and the area of a rectangle.	area = $A + B + C$ $= \left(\frac{1}{2}b \times h\right) + (b \times h) + \left(\frac{1}{2}b \times h\right)$ $= \left[\frac{1}{2} \times (2.0 \times 10^{-3}) \times 60\right] +$ $[(6.0 \times 10^{-3}) \times 60] +$ $\left[\frac{1}{2} \times (5.0 \times 10^{-3}) \times 60\right]$ $= 0.060 + 0.36 + 0.15$ $= 0.57$
The total impulse is equal to the area.	I = area = 0.57 kg m s^{-1}
Apply the sign and direction convention for motion in one dimension vertically.	I = 0.57 kg m s^{-1} up

Worked example: Try yourself 10.6.3

CALCULATING THE TOTAL IMPULSE FROM A CHANGING FORCE

A student records the force acting on a tennis ball as it bounces off a hard concrete floor over a period of time. The graph shows the forces acting on a ball during its collision with the concrete floor.



a Determine the force acting on the ball at a time of 4.0 milliseconds.

b Calculate the total impulse of the ball over the 13 millisecond period of time.

PHYSICS IN ACTION

Car safety

Designing a successful car is a complex task. A vehicle must be reliable, economical, powerful, visually appealing, secure and safe. Public perception of the relative importance of these issues varies. Magazines and newspapers concentrate on appearance, price and performance. The introduction of air-bag technology into most cars has altered the focus towards safety.

Vehicle safety is primarily about crash avoidance. Research shows potential accidents are avoided 99% of the time. The avoidance of accidents is mainly due to accident avoidance systems such as antilock brakes. When a collision does happen, passive safety features, such as the air bag, come into operation. Understanding the theory behind accidents involves primarily an understanding of impulse and force.

The airbag

The introduction of seatbelts allowed many more people to survive car accidents. However, many of these survivors sustained serious injuries. So, although seatbelts saved lives, there was also an increase in serious injuries. A further safety device was required to minimise these injuries.

The airbag in a car is designed to inflate within a few milliseconds of the occurrence of a collision to reduce secondary injuries during the collision. The airbag is designed to inflate only when the vehicle experiences an impact with a solid object at 18–20 km h^{-1} or more. The required deceleration must be high, or accidental nudges with another car would cause the airbag to inflate. The car's computer control makes a decision within a few milliseconds to detonate the gas cylinders that inflate the airbag. The propellant detonates and inflates the airbag while, according to Newton's first law, the driver continues to move towards the dashboard. As the driver continues forwards into the airbag, the bag deflates, allowing the body to slow down over a longer time than would otherwise be possible as it moves towards the dashboard (see Figure 10.6.3). The force is minimised so injury is reduced.

FIGURE 10.6.3 Airbags can prevent injuries by extending the period of time you take to stop.

Calculating exactly when the airbag should inflate, and for how long, is a difficult task. Many cars have been crash tested and the results painstakingly analysed. High-speed film demonstrates precisely why the airbag is so effective. During a collision the arms, legs and head of the occupants are restrained only by the joints and muscles. Enormous forces are involved because of the large deceleration. The shoulders and hips can, in most cases, sustain the large forces for the short duration. However, the neck is the weak link. Victims of road accidents regularly receive neck and spinal injuries. An airbag reduces the enormous forces the neck must withstand by extending the duration of the collision. This involves the direct application of the concept of impulse. A comparison of the forces applied to the occupant of a car with and without airbags is shown in Figure 10.6.4.

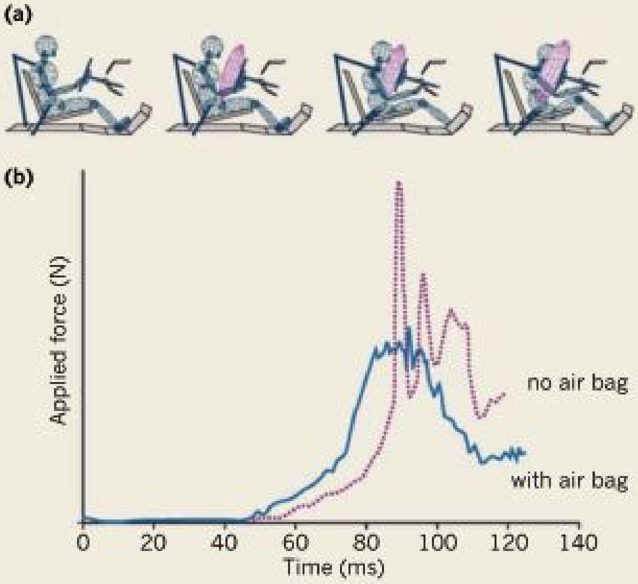


FIGURE 10.6.4 (a) The airbag extends the stopping time and distributes the force required to decelerate the mass of the driver or passenger over a larger area than a seatbelt. (b) The force withstood by the occupant of the car without an airbag is about double that felt with an airbag.

Airbags prevent the high forces caused by contact of the head with the steering wheel. The airbag ensures that the main thrust of the expansion is directed outwards instead of towards the driver. The airbag's deflation rate, governed by the size of the holes in the rear of the air bag, provides the optimum deceleration of the head for a large range of impact speeds.

The airbag is not the answer to all safety concerns associated with a collision, but it is one of many safety features that form a chain of defence in a collision.

10.6 Review

SUMMARY

- Newton's second law describes the relationship between impulse, force and the period of time: $I = F\Delta t$
- The same mass changing its velocity by the same amount will have a constant change in momentum or impulse.
- The faster a mass changes its velocity, the greater the force required to change the velocity in that period of time.
- The slower a mass changes its velocity, the smaller the force required to change the velocity in that period of time.
- Forces can change during a collision.
- The impulse over a period of time can be found by calculating the area under the line on a force versus time graph.
- The period of time is the cause of the difference between the force provided by two different surfaces during a collision. Hard surfaces result in a short time to stop and soft surfaces result in a longer time to stop.
- The effect of the period of time on the force is dramatic. A shorter time means a greater force, while a longer time means a much smaller force.

KEY QUESTIONS

1 A 45.0 kg mass changes its velocity from 2.45 m s^{-1} east to 12.5 m s^{-1} east in a period of 3.50 s.
   - **a** Calculate the change in momentum of the mass.
   - **b** Calculate the impulse of the mass.
   - **c** Calculate the force that causes the impulse of the mass.

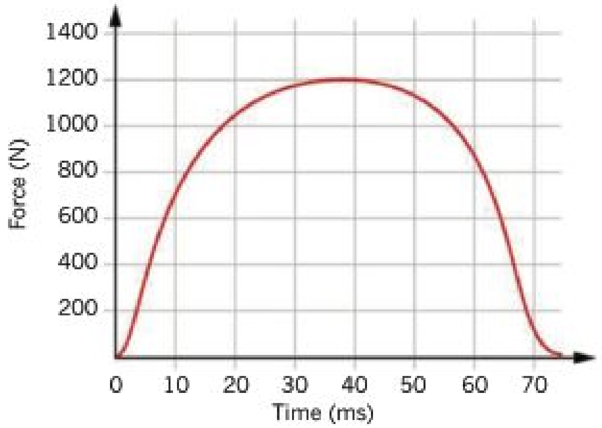
2 Using the concept of impulse, explain how airbags can reduce injuries during a collision.

3 A student catches a 75.0 g cricket ball with 'hard hands'. Just before the student catches the ball, its velocity is 15.6 m s^{-1} west. With hard hands, the velocity of the ball drops to zero in just 0.100 seconds.
   - **a** Calculate the change in momentum of the cricket ball.
   - **b** Calculate the impulse of the cricket ball.
   - **c** Calculate the average force on the cricket ball.

4 The student from question 3 now catches the same 75.0 g cricket ball, but this time with 'soft hands'. Just before the student catches the ball, its velocity is 15.6 m s^{-1} west. With soft hands, the velocity of the ball drops to zero in 0.300 seconds. Calculate the average force on the cricket ball.

5 A 200 g cricket ball (at rest) is struck by a cricket bat. The ball and bat are in contact for 0.05 s, during which time the ball is accelerated to a speed of 45 m s^{-1}.
   - **a** What is the magnitude of the impulse the ball experiences?
   - **b** What is the net average force acting on the ball during the contact time?
   - **c** What is the net average force acting on the bat during the contact time?

6 The following graph shows the net vertical force generated as an athlete's foot strikes an asphalt running track.



   - **a** Estimate the maximum force acting on the athlete's foot during the contact time.
   - **b** Estimate the total impulse during the contact time.

7 A 25 g arrow buries its head 2 cm into a target on striking it. The arrow was travelling at 50 m s^{-1} just before impact.
   - **a** What change in momentum does the arrow experience as it comes to rest?
   - **b** What is the impulse experienced by the arrow?
   - **c** What is the average force that acts on the arrow during the period of deceleration after it hits the target?

8 Crash helmets are designed to reduce the force of impact on the head during a collision.
   - **a** Explain how their design reduces the net force on the head.
   - **b** Would a rigid 'shell' be as successful? Explain.

Chapter review

10

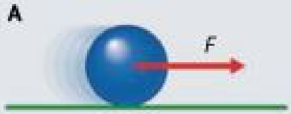
KEY TERMS

conserved
contact forces
force
impulse
inertia
momentum
net force
newton
Newton's first law
Newton's second law
Newton's third law
non-contact forces

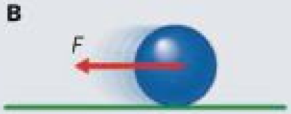
1 A student is travelling to school on a train. When the train starts moving, she notices that passengers tend to lurch towards the back of the train before grabbing a handrail to stop themselves from falling. Has a force acted to push the passengers backwards? Justify your answer.

2 A bowling ball rolls along a smooth wooden floor at constant velocity. Ignoring the effects of friction and air resistance, which of the following diagrams correctly indicates the forces acting on the ball?

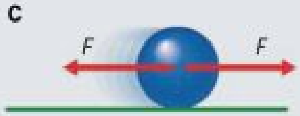
A



B



C



D

3 Calculate the mass of an object if it accelerates at 9.20 m s^{-2} east when a force of 352 N east acts on it.

4 Calculate the acceleration of a 657 kg motorbike when a net force of 3550 N north acts on it.

The following information relates to questions 5–8.

Lachy is riding his bike and producing a forwards force of 150 N. The combined mass of Lachy and the bike is 100 kg.

5 If there is no friction or air resistance, what is the magnitude of the acceleration of Lachy and the bike?

6 If friction opposes the bike's motion with a force of 45.0 N, what is the magnitude of the acceleration of the bike?

7 What must be the magnitude of the force of friction if Lachy's acceleration is 0.600 m s^{-2}?

8 Lachy now carries an additional mass of 25.0 kg due to his school bag. What must be the new forwards force he produces in order to accelerate at 0.800 m s^{-2} if friction opposes the motion with a force of 30.0 N?

9 Calculate the final velocity of a 14.0 kg remote-controlled car moving at 3.75 m s^{-1} east, when a net force of 62.0 N west acts on it for 2.00 seconds.

10 A 65.0 kg student is standing on a 3.50 kg skateboard at rest when he steps off the board and exerts a horizontal force of 75.0 N south on the board. What force does the board exert on the student?

11 Calculate the change in momentum of a 155 kg rock whose velocity has changed from 6.50 m s^{-1} east to 3.25 m s^{-1} east in a period of 8.50 s.

12 Calculate the change in momentum of a 25.5 kg robot whose velocity changes from 6.40 m s^{-1} forwards to 2.25 m s^{-1} backwards.

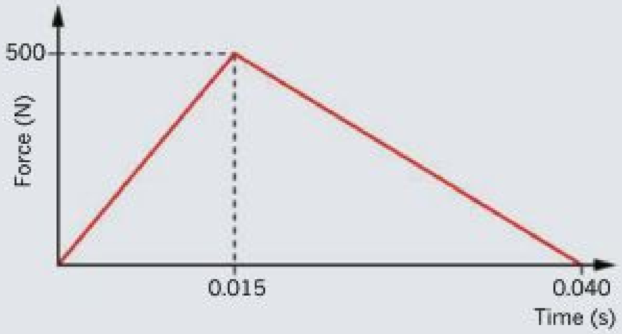
13 An astronaut in a protective suit has a total mass of 154 kg and throws a 40.0 kg toolbox away from the space station. The astronaut and toolbox are initially stationary. After being thrown, the toolbox moves at 2.15 m s^{-1}. Calculate the velocity of the astronaut just after throwing the toolbox.

14 A 75.0 kg netball player is moving at 4.00 m s^{-1} west changes direction during a game to 5.00 m s^{-1} north. Calculate the change in momentum of the player over the period of the change.

15 An athlete catches a 300 g volleyball by relaxing her elbows and wrists and 'giving' with the ball. Just before the athlete catches the ball, its velocity Is 5.60 m s^{-1} west. With soft hands, the velocity of the ball drops to zero in 1.00 second. Calculate the average force exerted by the athlete on the volleyball.

The following information relates to questions 16–18.

Jordy is playing softball and hits a ball with her softball bat. The force versus time graph for this interaction is shown below. The ball has a mass of 170 g.



16 Determine the magnitude of the change in momentum of the ball.

17 Determine the magnitude of the change in momentum of the bat.

18 Determine the magnitude of the change in velocity of the ball.

CHAPTER

11

Equilibrium of forces

In the design of buildings and other structures, engineers and architects must use their physics knowledge to determine the forces that act within the structures that they create.

This chapter will cover the concept of equilibrium, which describes the situation in which forces and torques are balanced. If there is equilibrium of translational (linear) forces then there will be no net translational forces, and an object will not begin to move. If there is equilibrium of torque, then the object will not rotate.

Key knowledge

By the end of this chapter you will have studied equilibrium and forces acting on a structure. You will know how to calculate and balance those forces, and will be able to:

- calculate torque: $\tau = r_{\perp}F$
- investigate and analyse theoretically and practically translational forces and torques in simple structures that are in rotational equilibrium.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

11.1 Torque

Many situations involve objects that rotate about a pivot point, such as closing a door, using a spanner, or turning a steering wheel. In these situations, a force acts to provide a turning effect or a **torque** (τ). Newton's laws use the concept of a force to help understand changes in the linear (straight line) motion of an object. The concept of torque is used in exactly the same way to explain a change in the rotational (turning) motion of an object.

TORQUE

FIGURE 11.1.1 Applying a torque to a steering wheel will cause it to turn.

Consider the steering wheel in Figure 11.1.1. When a turning effect is applied to the steering wheel there are a number of factors that must work together, causing it to turn. For all turning objects, there must be a **pivot point**, around which the object will rotate. There must be a force (F) applied to the object in such a way as to cause the object to rotate. This means that the force applied must not be aligned with the pivot point. There must be some distance between the **line of action of the force** (an imaginary line through the force vector) and the pivot point.

EXTENSION

Torque and moments

The terms torque (τ) and moment of a force (M), which is usually shortened to moment, are terms that are used interchangeably in Physics at the high-school level.

The difference between the terms is in how they are used. Typically, torque is used for dynamic problems where there is an angular acceleration, which means the speed at which an object is rotating is changing. This angular acceleration is caused by a net torque with a non-zero value. Moments are typically used in static problems in which there is no rotation. In these cases there is no rotation because the moments, which are often created by reaction forces or internal forces in an object, are balanced so that there is no angular acceleration.

For example, the term torque might be used to describe the effect of the force applied by an axle to a car's wheel to increase its speed of rotation. The term moment might be used to describe the effect of the reaction forces acting on a diving board. When you stand on the tip of a diving board, the reaction forces at the other end must not only push upwards (to counteract your weight) but they must create a moment to prevent the diving board from rotating.

Both torque and moment are calculated using the same equation. For the purposes of this chapter we can consider the two terms as interchangeable.

Force and the pivot point

When analysing a rotating system, the position of the pivot point or **axis of rotation** is an important consideration. A wheel, for example, moves in a circular path around its axle. An imaginary line along the length of the axle is called the axis of rotation and is shown in Figure 11.1.2.

The pivot point is the point on a two-dimensional representation of the object through which the axis of rotation passes. As an example, the pivot point of a wheel is shown in Figure 11.1.3.

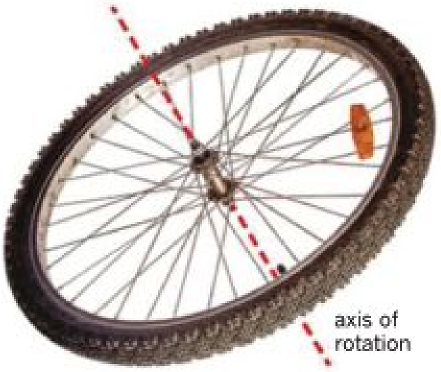


FIGURE 11.1.2 The axis of rotation.

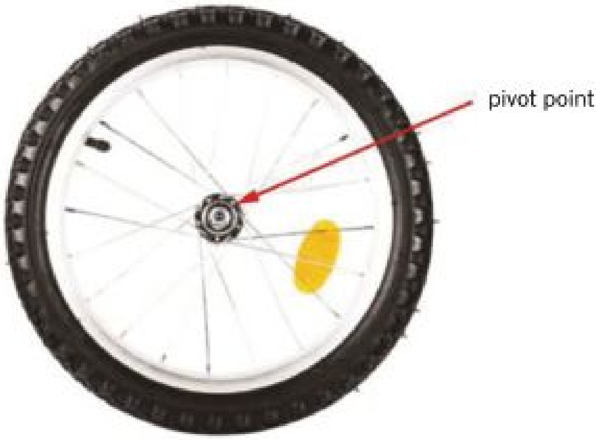


FIGURE 11.1.3 The pivot point.

A force applied directly towards or directly away from the pivot point of the wheel will not create a turning effect on the wheel. So, for the example in Figure 11.1.4, if the force acted along the line labelled 'line of action', the wheel would not turn.

FIGURE 11.1.4 When the line of action of the force passes through the pivot point, the wheel will not turn.

Torque can be achieved by applying a force on the wheel where the line of action of the force does not pass through the axis of rotation or the pivot point. The maximum effect is achieved when the force applied is at 90° to a line drawn from the pivot point to the point of application (the point at which the force is applied). This is shown in Figure 11.1.5.

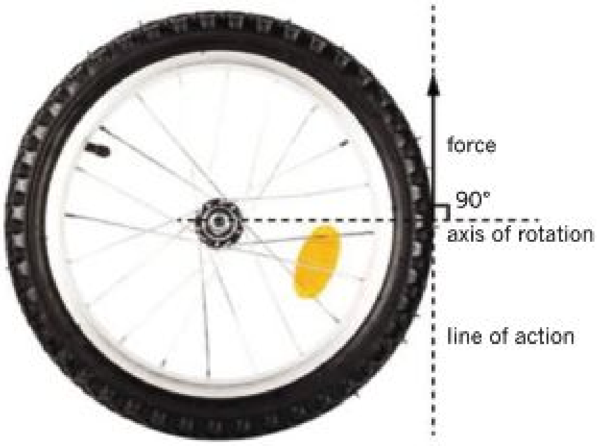


FIGURE 11.1.5 Maximum torque occurs when the force applied is perpendicular (at 90°) to a line drawn from the pivot point to the point of application.

Magnitude of the force and torque

The torque (τ) on an object is directly proportional to the magnitude of the force (F). If all other things are equal, a larger force will result in a larger torque. This is illustrated in Figure 11.1.6.

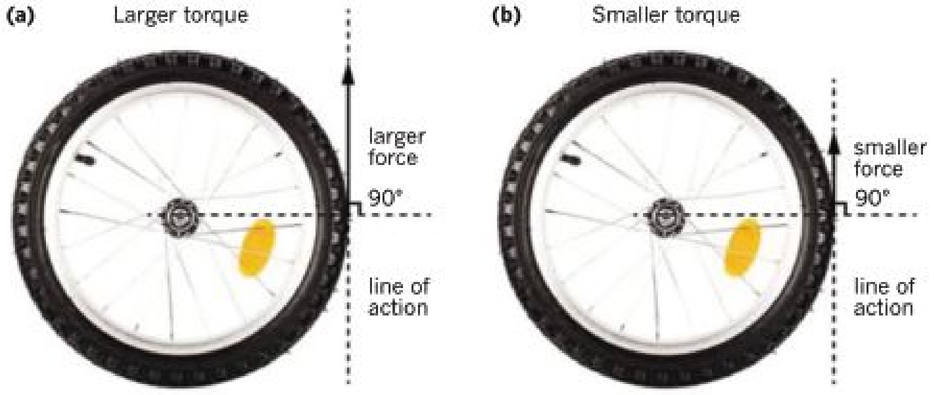


FIGURE 11.1.6 The magnitude of the force affects the torque on an object. The wheel in (a) will experience a larger torque than the wheel in (b).

Distance from the pivot point and torque

The amount of torque created is directly proportional to the perpendicular distance between the pivot point and the line of action of the force. This perpendicular distance is called the **force arm**. The force arm is given the symbol $r_\perp$ and is shown in Figure 11.1.7. Given that everything else is constant, then the larger the force arm or perpendicular distance ($r_\perp$), the larger the torque (τ).

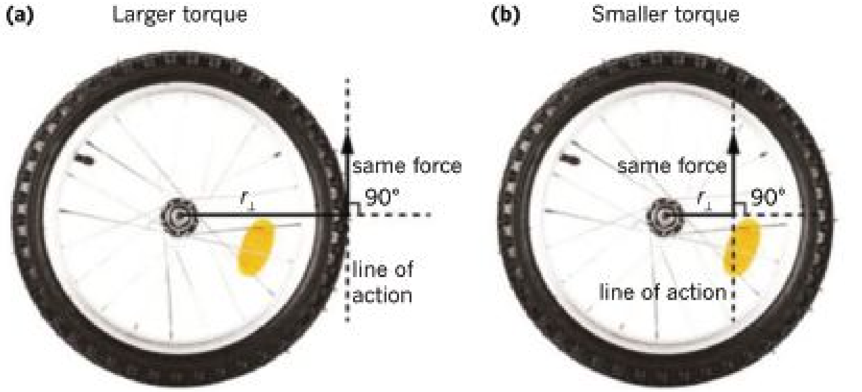


FIGURE 11.1.7 The perpendicular distance from the pivot point to the line of action of the force affects the torque on an object. The wheel in (a) will experience a larger torque than the wheel in (b).

It was stated previously that maximum torque occurs when the line of action of the force is perpendicular to a line drawn from the pivot point to the point of application. You can now use the concept of a force arm to understand this. The force arm is maximised when the line of action is perpendicular to the line between the pivot point and the point of application, and therefore the torque is maximised (see Figure 11.1.8).

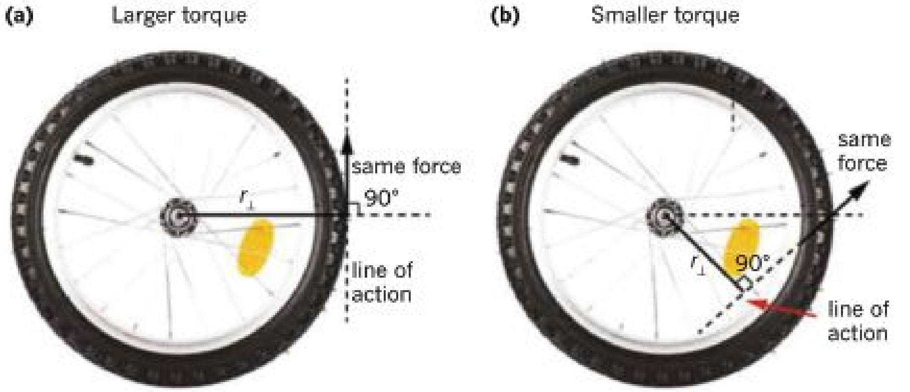


FIGURE 11.1.8 The direction of the applied force affects the size of the force arm, and therefore the torque on the object. The wheel in (a) will experience a larger torque than the wheel in (b).

The torque equation

The magnitude of the torque (τ) increases or decreases as the force (F) increases or decreases. The magnitude of the torque also increases or decreases as the force arm or the perpendicular distance from the pivot to the line of action of the force ($r_\perp$) increases or decreases.

The formula for calculating torque is:

$\tau = r_\perp F$

where τ is the torque (N m)

$r_\perp$ is the force arm (m)

F is the force (N).

Torque is a vector quantity. This enables us to distinguish its unit from the joule, the scalar unit for work and energy, which can also be written as N m. A rotating body rotates either clockwise or anticlockwise. A clockwise rotation is considered to be positive and an anticlockwise rotation to be negative. This convention is useful when more than one torque is acting on a body and the net torque has to be found.

Worked example 11.1.1

CALCULATING TORQUE

A bus driver applies a force of 45.0 N on the steering wheel of a bus as it turns a right-hand corner. The radius of the steering wheel is 30.0 cm. If the force is applied at 90° to the radius, calculate the torque on the steering wheel.	
Thinking	**Working**
Identify the variables involved and state them in their standard form.	$\tau = ?$ $r_\perp = 0.300$ m $F = 45.0$ N
Apply the equation for torque.	$\tau = r_\perp F$ $= 0.300 \times 45.0$ $= 13.5$ N m
State the answer with the appropriate direction.	$\tau = 13.5$ N m clockwise

Worked example: Try yourself 11.1.1

CALCULATING TORQUE

A force of 255 N is required to apply a torque on a sports car steering wheel as it turns left. The force is applied at 90° to the 15.5 cm radius of the steering wheel. Calculate the torque on the steering wheel.

Torque on different objects

Torque doesn't need to be acting on circular objects only. Any object can rotate about a point if a force is applied where the line of action of the force is not acting through the pivot point.

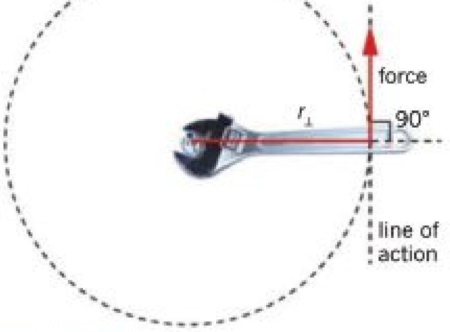


FIGURE 11.1.9 Although the adjustable spanner is not a wheel or circle, torque can still be applied to the nut.

Spanners, like the one in Figure 11.1.9, apply a torque to a nut or bolt: the pivot point is the bolt and a force is applied at right angles to the spanner.

The reason a spanner is an effective hand tool is because it increases the force arm when turning a nut. If you try unscrewing a nut with your hands you will probably find that you are unable to provide enough force to create the torque required to turn the bolt. Longer spanners can apply a greater torque on a nut than shorter spanners. Some wheelnut spanners, like the one in Figure 11.1.10, have handles which can extend so the force arm can be increased. This provides extra torque for loosening very tight nuts or tightening the nuts to the correct torque.

FIGURE 11.1.10 Removing a tyre with an extended handle spanner will increase the torque on the nut.

Worked example 11.1.2

CALCULATING PERPENDICULAR DISTANCE

A car driver can apply a maximum force of 845 N on a wheelnut spanner that is adjustable up to 30.0 cm in length. The force is applied at 90° to the spanner. If the wheelnuts need a torque of 224 N m to remove them, what is the minimum length of the adjustable spanner so that the nuts can be loosened? State whether the spanner is long enough.

Thinking	Working
Identify the variables involved and state them in their standard form.	$\tau = 224$ N m $r_\perp = ?$ $F = 845$ N
Apply the equation for torque. Rearrange if necessary.	$\tau = r_\perp F$ $r_\perp = \frac{\tau}{F}$ $= \frac{224}{845}$ $= 0.265$ m
State the answer with the appropriate units.	$r_\perp = 26.5$ cm
Compare the answer with the length of the spanner and state whether it is or isn't appropriate for this task.	As the spanner can be extended to 30.0 cm this is long enough to provide the minimum perpendicular distance of 26.5 cm. So the spanner is long enough.

Worked example: Try yourself 11.1.2

CALCULATING PERPENDICULAR DISTANCE

A truck driver can apply a maximum force of 1022 N on a large truck wheelnut spanner that has a length of 80.0 cm. The force is applied at 90° to the spanner. If the truck's wheelnuts need a torque of 635 N m to make them secure, determine if the length of this spanner is sufficient for the job.

Doors are also good examples of torque in action, with the hinges forming the axis of rotation. If force is applied to the handle and the line of action of the force is perpendicular to the door, then the distance between the hinge and the handle represents the force arm. This is shown in Figure 11.1.11.

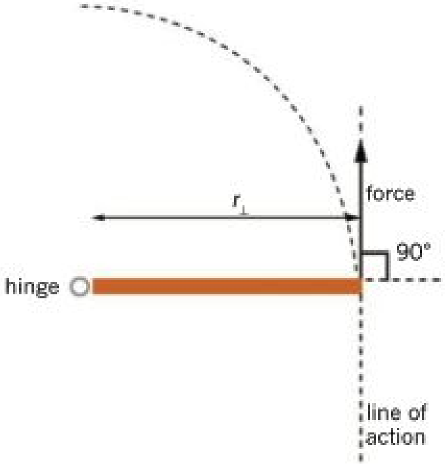


FIGURE 11.1.11 A door can have a torque applied to it, as long as the line of action of the force is not through the axis of rotation.

NON-PERPENDICULAR CALCULATIONS OF TORQUE

When the force causing a torque acts along a line that is at an angle other than 90° to an object (such as the door in Figure 11.1.12) then the torque is reduced. In these circumstances, we can calculate the torque by two approaches: either finding the component of the force acting perpendicular to the door, or by finding the perpendicular distance from the pivot point to the line of action of the force.

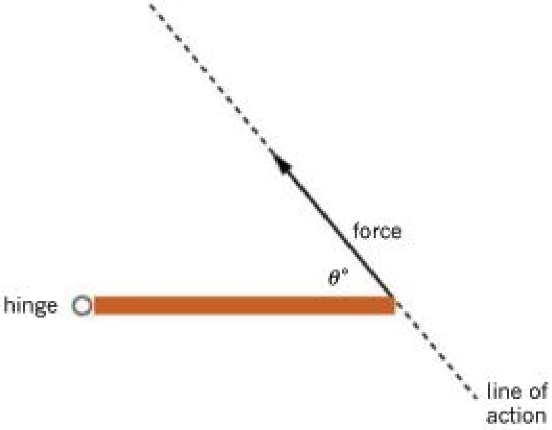


FIGURE 11.1.12 When the force causing a torque is not perpendicular to a door, the torque is reduced.

Recall that the formula for torque (τ) on an object is:

$$\tau = r_\perp F$$

This equation calculates the torque (τ) when the force (F) and the distance from the pivot to the line of action of the force (r) are perpendicular to each other. It really doesn't matter whether the radius is perpendicular to the line of action of the force, or if the force is perpendicular to the radius. The result is the same either way. That is, $\tau = r_\perp F$ and also $\tau = rF_\perp$.

Calculating torque using perpendicular force

The components of any force can be calculated using trigonometry. To find the component of the force that is perpendicular to a door, for example, use the magnitude of the force and the angle between the door and the line of action of the force. This is shown in Figure 11.1.13.

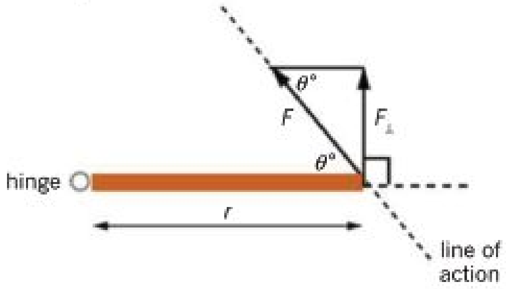


FIGURE 11.1.13 Finding the component of the force that is perpendicular to a door using the magnitude of the force and the angle between the door and the line of action of the force.

In this case, $\tau = rF_{\perp}$ and $F_{\perp} = F \sin \theta$.

This combines to give $\tau = r F \sin \theta$.

Your strategy for solving questions of this type may be to calculate the perpendicular component of force and then apply the torque equation, or to use the combined equation. To begin with it is recommended that you calculate the perpendicular component and then use the torque equation. When you have gained confidence with that strategy, try using the combined equation.

Worked example 11.1.3

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF FORCE

A student uses a 42.0 cm long adjustable spanner to loosen a nut on her bike. She applies a force of 65.0 N at an angle of 68.0° to the spanner.

Calculate the anticlockwise torque that the student applies to the nut.

Thinking	Working
Use the trigonometric relationship $F_{\perp} = F \sin \theta$ to determine the force perpendicular to the spanner.	$F_{\perp} = F \sin \theta$ $= 65.0 \sin 68.0$ $= 60.3$ N
Convert the variables to their standard units.	$r = 42.0$ cm $= 0.420$ m
Apply the equation for torque: $\tau = rF_{\perp}$	$\tau = rF_{\perp}$ $= 0.420 \times 60.3$ $= 25.3$
State the answer with the appropriate units.	$\tau = 25.3$ N m anticlockwise

Worked example: Try yourself 11.1.3

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF FORCE

A mechanic uses a 17.0 cm-long spanner to tighten a nut on a winch. He applies a force of 104 N at an angle of 75.0° to the spanner.

Calculate the magnitude of the torque that the mechanic applies to the nut.
Give your answer to three significant figures.

Calculating torque using perpendicular radius

The component of any distance can be calculated using either Pythagoras' theorem or trigonometry. To find the component of a length that is perpendicular to the line of action of the force acting on a door, construct a line from the pivot point to the line of action of the force so that it intersects the line of action at right angles. An example is shown in Figure 11.1.14.

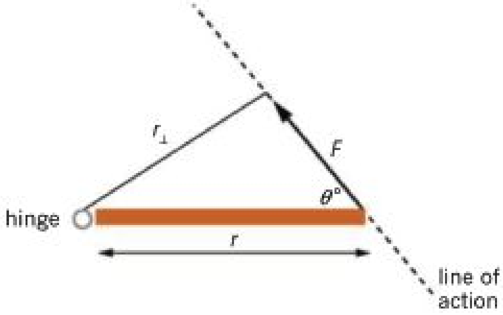


FIGURE 11.1.14 Determining the components of a distance.

In this case:

$\tau = r_\perp F$ and $r_\perp = r \sin \theta$

combine to give:

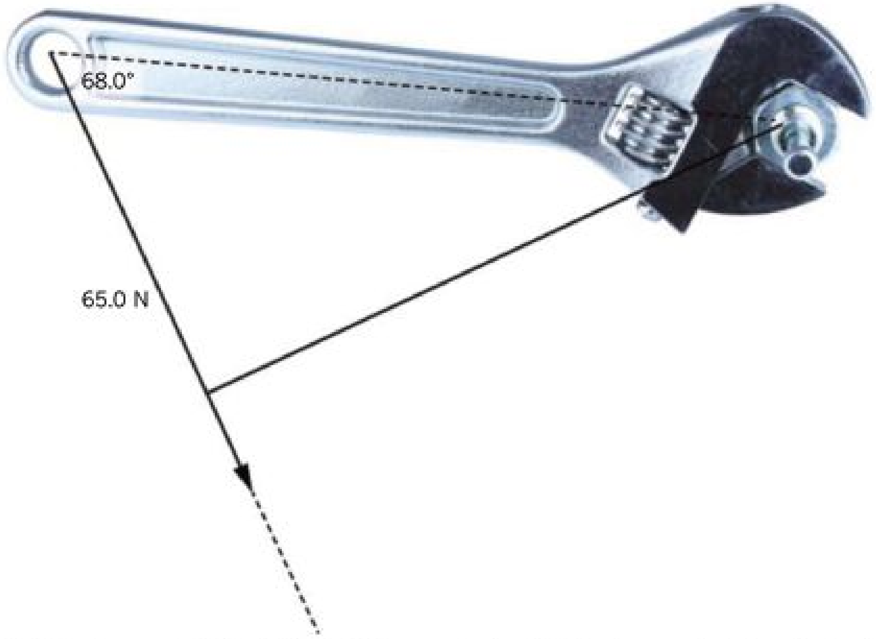
$\tau = r \sin \theta F$

The equation $\tau = rF \sin \theta$ is identical to $\tau = r \sin \theta F$, so either method would be appropriate for calculating the torque on an object when the force is not at right angles to the object. In either method, the component of the distance or the component of the force is always going to be less than the total distance or the total force itself. This will result in a smaller torque being applied to the object. The maximum torque will always be when the line of action of the force is perpendicular to the distance from the pivot point to the point of application.

Worked example 11.1.4

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF DISTANCE

A student uses a 42.0 cm long adjustable spanner to loosen a nut on her bike. She applies a force of 65.0 N at an angle of 68.0° to the spanner.



In Worked example 11.1.3 (p. 382), you calculated the torque that the student applies to the nut using the perpendicular force. Now, using the perpendicular distance, calculate the anticlockwise torque that the student applies to the nut.

Thinking	Working
Convert variables to their standard units.	$r = 42.0$ cm $= 0.420$ m
Use the trigonometric relationship $r_\perp = r \sin\theta$ to determine the perpendicular distance from the pivot point to the line of action of the force.	$r_\perp = r \sin\theta$ $= 0.420 \sin 68.0$ $= 0.389$ m
Apply the equation for torque: $\tau = r_\perp F$	$\tau = r_\perp F$ $= 0.389 \times 65.0$ $= 25.3$
State the answer with the appropriate units. Note that this is the same answer as in the previous Worked example.	$\tau = 25.3$ N m anticlockwise

Worked example: Try yourself 11.1.4

CALCULATING TORQUE FROM THE PERPENDICULAR COMPONENT OF DISTANCE

A mechanic uses a 17.0 cm-long spanner to tighten a nut on a winch. He applies a force of 104 N at an angle of 75.0° to the spanner.

Using the perpendicular distance, calculate the magnitude of the torque that the mechanic applies to the nut.

Give your answer to three significant figures.

PHYSICS IN ACTION

The torque wrench

The extent to which a nut or bolt is tightened can be critical to the safe operation of machinery or motors. If a nut or bolt is too loose then it could fall out. If it is too tight then it could either distort the part or the bolt could break off. Both of these situations could require expensive repairs. To avoid nuts and bolts being too loose or too tight, manufacturers use different tools and methods to estimate the amount of torque required to tighten a nut or bolt to the correct tightness. Some examples of these tools are shown in Figure 11.1.15.

The beam wrench is the simplest type of toque wrench. It has a flexible lever arm with a bar and scale, separating the wrench head and handle. When torque is applied, a pointer on the scale moves to indicate the amount of torque being applied in newton metres (N m).

The clicking type torque wrench can be set to apply a fixed amount of torque. When the required amount of torque has been achieved, the wrench 'clicks' and releases itself, preventing any further tightening from being applied.

More recently, electronic torque wrenches have been developed. The signal generated is converted to a torque reading (N m) and is shown on the digital readout screen. Measurements can also be stored within the instrument's memory and transferred to a computer.

FIGURE 11.1.15 Three types of wrenches commonly used to measure the torque applied to a nut or bolt: (a) beam torque wrench, (b) click-type torque wrench, and (c) digital torque wrench.

11.1 Review

SUMMARY

- Torque is a measurement of the tendency of a force to cause an object to rotate around an axis.
- The formula for calculating torque is $\tau = r_{\perp}F$.
- Torque occurs when the acting force is not applied directly through the pivot point of the object.
- Maximum torque occurs when the acting force applied is perpendicular to a line drawn from the pivot point to the point of application.
- The larger the force acting on the object, the larger the torque will be.
- The longer the force arm, the greater the torque will be.
- If torque is generated by an acting force that is not perpendicular to the lever arm of the object, then either:
  - the component force perpendicular to the length of the object is used to calculate torque: $\tau = rF_{\perp}$

    or
  - the distance from the pivot point perpendicular to the line of action of the force is used to calculate torque: $\tau = r_{\perp}F$.
- Both strategies for determining torque from non-perpendicular situations equate to $\tau = rF \sin \theta$.

KEY QUESTIONS

Use $g = 9.80$ N kg^{-1} when answering these questions.

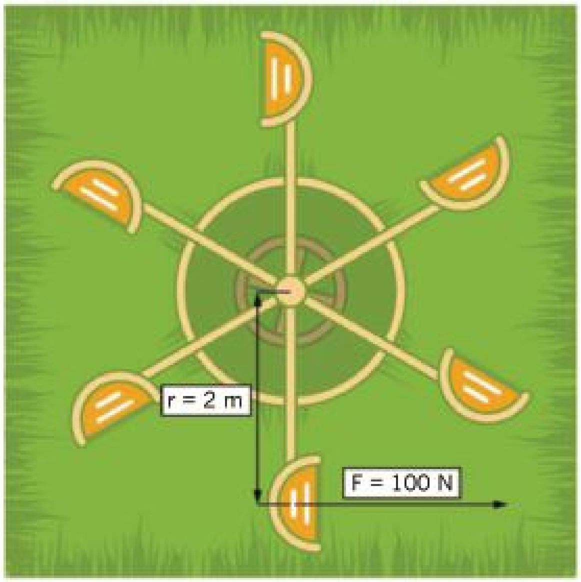
1 Use the concept of torque to explain the following.

a It is easier to open a heavy door by pushing it at the handle rather than in the middle of the door.

b It is possible to move very heavy rocks in the garden by using a long crowbar.

2 Calculate the torque exerted on the roundabouts shown below. Include the direction where appropriate.

a



b

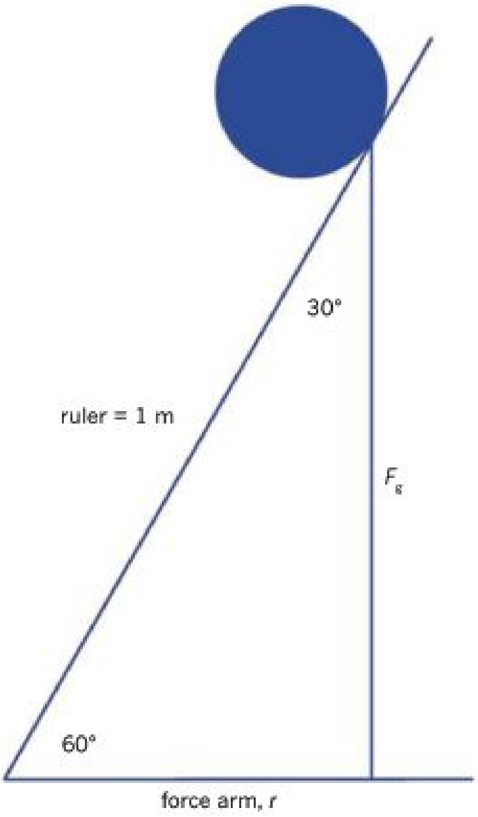
3 The magnitude of the torque required to tighten a bicycle wheel is 15 N m. Calculate the force arm required if a 30 N force is used.

4 A student pushes a heavy door at a point that is 50 cm from the hinges such that it creates a torque of 9 N m. With what magnitude of force does the student push? Assume the student pushes perpendicular to the surface of the door.

5 A spanner with a length of 40 cm is used to tighten a nut on a car wheel. If the magnitude of force applied is 225 N, calculate the maximum torque that can be created on this wheelnut.

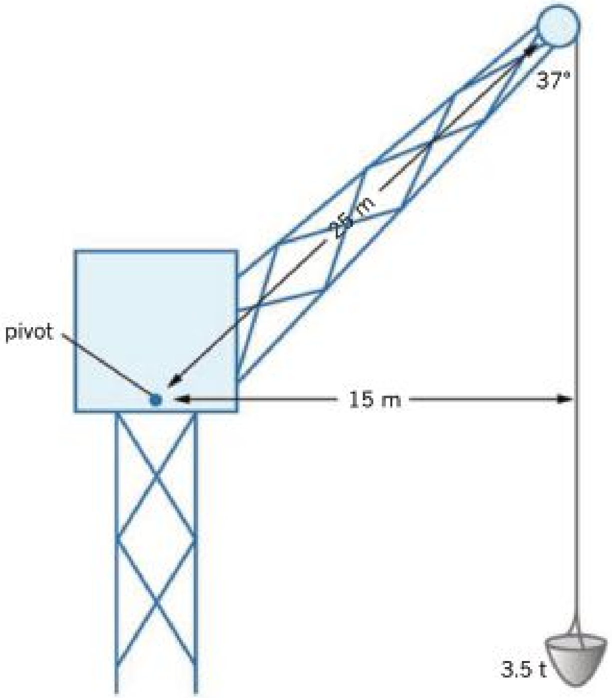
6 Nikki is investigating torque using a metre rule and a 1.0 kg mass. She uses a rubber band to attach the mass to the ruler. Nikki first holds the ruler at one end so that it is horizontal, with the mass at the 50 cm mark.

- **a** What is the size of the torque that is acting?
- **b** She now moves the mass so that it is right at the far end of the ruler. How much torque is acting now?
- **c** Finally, she lifts the ruler so that it makes an angle of 60° to the horizontal. What is the size of the torque now?



7 A mechanic uses a spanner of length 30 cm to tighten a bolt head. The mechanic applies a force of 300 N at an angle of 30° to the length of the spanner. Calculate the torque created.

8 A crane is being used to lift a skip of concrete with a total mass of 3.5 tonnes (3500 kg). The lever arm of the crane is 25 m long and makes an angle of 37° with the vertical as shown in the diagram. Ignore the mass of the cable when answering these questions.



- **a** What is the total weight of the skip?
- **b** The skip is lifted so that it is near the top of the crane. Does the torque around the pivot created by this load increase, decrease or remain the same as the load is lifted?
- **c** Calculate the magnitude and direction of the torque about the pivot that the skip exerts on the crane when the skip is at the highest point.

11.2 Translational equilibrium

Newton's first law states that an object will continue with its motion unless acted upon by an external unbalanced force. An object's velocity won't change when the forces acting on it are balanced. When the forces are balanced, the forces are said to be in translational equilibrium.

An example of translation equilibrium occurs at the beginning of a game of tug-o-war, like that in Figure 11.2.1. Both teams take the strain and neither team moves. Winning a tug-o-war game involves one team applying a greater force so that there is a net force on the rope, causing the rope and teams to accelerate in the winning team's direction. When the rope and the teams are moving at a constant velocity, then an equilibrium of forces exists once again.

FIGURE 11.2.1 When a tug-o-war starts, there is an equilibrium of forces as both teams take the strain.

TRANSLATIONAL EQUILIBRIUM

A **translational equilibrium** of forces occurs when the sum of the forces acting on an object add to give a zero resultant force or zero net translational force. As a net translational force causes acceleration in one direction, a zero net translational force causes no acceleration of the object. This condition is the defining aspect of a translational equilibrium of forces.

When analysing a situation involving more than one force acting on an object, translational equilibrium will exist if the sum of the forces is equal to zero: $\Sigma F = 0$. The sum of forces is commonly referred to as the net force, and so the condition for translational equilibrium can be expressed as: $F_{\text{net}} = 0$.

Vector diagrams of an equilibrium of forces in one dimension

Vector diagrams can be drawn to represent the forces acting on an object when the forces are acting in one dimension. For example, if three people are pulling to the right and three people are pulling to the left in a game of tug-o-war as shown in Figure 11.2.2, then the forces are all in one dimension—left and right. These forces are added using a vector diagram by drawing all the forces from each person head to tail, as described in Chapter 8. If the tug-o-war is in translational equilibrium, then all the forces should add to give a zero net force.

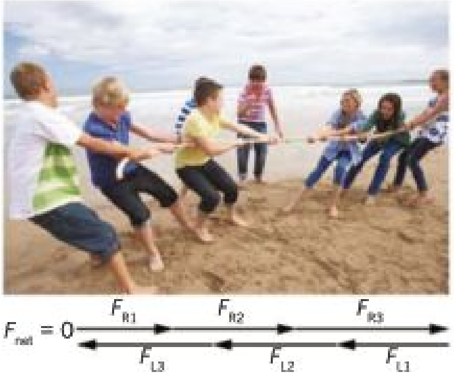


FIGURE 11.2.2 Tug-o-war with a vector addition diagram showing a net force of zero, indicating an equilibrium of forces exists.

Calculating an equilibrium of forces in one dimension

To calculate whether a situation is in translational equilibrium or not, a sign convention is used to represent the direction of the force vectors. Typically, in the *x*-dimension left is negative and right is positive; similarly in the *y*-dimension up is positive and down is negative. In the *z*-dimension forwards is positive and backwards is negative. By applying a sign convention to the forces acting on an object, the addition of those forces, with their signs, will give a zero answer if the situation is in translational equilibrium.

$$F_{\text{net, left–right}} = 0 \text{ or } F_{\text{net}, x} = 0$$

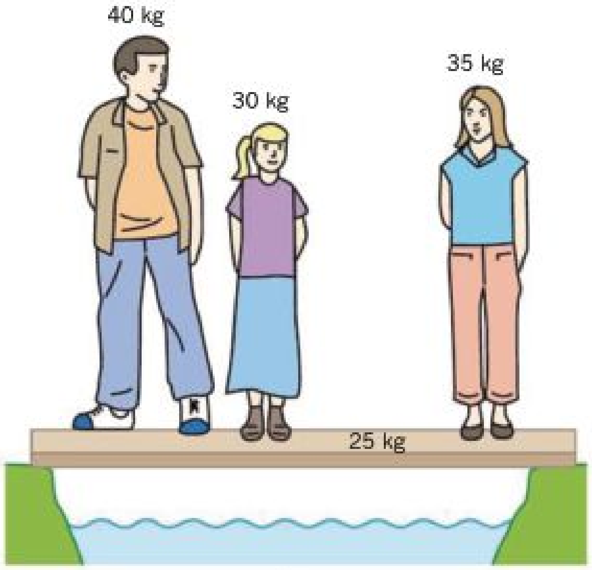
$$F_{\text{net, up–down}} = 0 \text{ or } F_{\text{net}, y} = 0$$

$$F_{\text{net, forwards–backwards}} = 0 \text{ or } F_{\text{net}, z} = 0$$

Worked example 11.2.1

CALCULATING TRANSLATIONAL EQUILIBRIUM IN ONE DIMENSION

Three children are standing on a plank that is bridging a small stream. The plank is supported at each end by the ground. The plank has a mass of 25.0 kg and the children have masses of 40.0 kg, 30.0 kg and 35.0 kg. There is an upwards force of the left bank on the plank of 700 N. If the plank is in translational equilibrium then calculate the force of the right bank on the plank. Use $g = 9.80\ \text{N kg}^{-1}$ when answering this question.

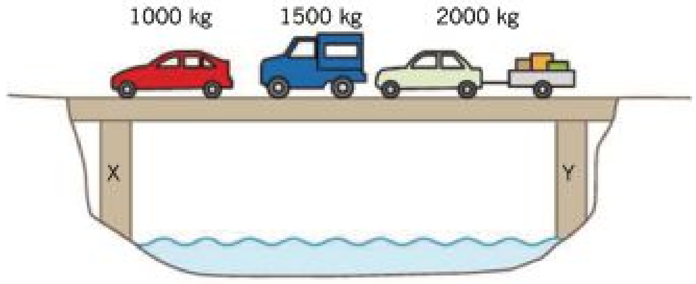


Thinking	**Working**
Identify the variables involved and state them with their directions in their standard form.	$m_1 = 40.0$ kg $m_2 = 30.0$ kg $m_3 = 35.0$ kg $m_P = 25.0$ kg $F_{LB} = 700$ N up $g = 9.80\ \text{N kg}^{-1}$ down
Apply a sign convention to the vector data.	$F_{LB} = +700$ N $g = -9.80\ \text{N kg}^{-1}$
Identify the object that is in translational equilibrium. This is the object on which all the forces are acting.	The object experiencing translational equilibrium is the plank.
Apply the equation for translational equilibrium in one dimension.	$F_{\text{net},\,y} = 0$
Expand the equation to include each of the forces acting on the plank.	$F_1 + F_2 + F_3 + F_P + F_{LB} + F_{RB} = 0$ $m_1g + m_2g + m_3g + m_Pg + F_{LB} + F_{RB} = 0$
Substitute the data into the equation and solve for the unknown.	$(40.0 \times -9.80) + (30.0 \times -9.80) + (35.0 \times -9.80) + (25.0 \times -9.80) + 700 + F_{RB} = 0$ $-392 - 294 - 343 - 245 + 700 + F_{RB} = 0$ $-574 + F_{RB} = 0$ $F_{RB} = 574$ N
State the answer with the appropriate direction.	$F_{RB} = 574$ N up

Worked example: Try yourself 11.2.1

CALCULATING TRANSLATIONAL EQUILIBRIUM IN ONE DIMENSION

Three cars are parked on a beam bridge that has a mass of 500 kg. The left pillar (labelled X) applies a force of 2.00×10^4 N upwards. If the situation is in translational equilibrium then calculate the force provided by the right-hand pillar (labelled Y). Use $g = 9.80$ N kg^{-1} when answering this question.



If the forces involved in the equilibrium situation are in two dimensions, there are two strategies for determining if the sum of the forces is zero.

Vector diagrams of an equilibrium of forces

In any vector addition diagram, the individual vectors are added head to tail. The net (resultant) vector is from the tail of the first vector to the head of the last vector. In a situation where the forces are in equilibrium, the vector addition diagram should end up with the head of the last vector finishing at the tail of the first vector. This means that the vector addition diagram ends up in a closed loop and therefore there is no net force. Figure 11.2.3 shows vector diagrams where (a) the net force (F_{net}) is not zero and (b) where F_{net} is zero.

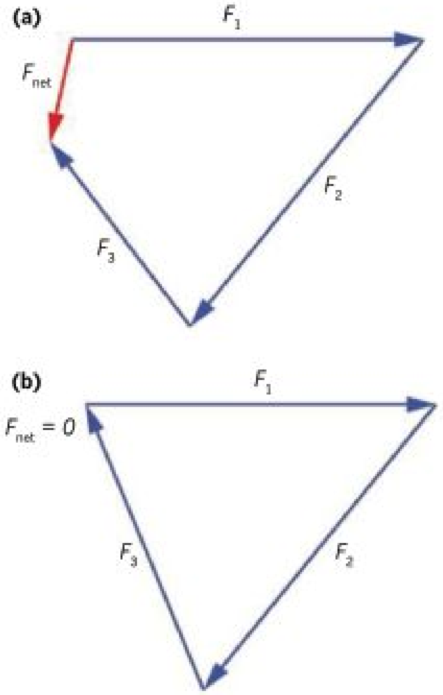


FIGURE 11.2.3 (a) A vector diagram showing three vectors added head to tail, and the resultant net force vector F_{net} in red. (b) A closed loop vector addition diagram, showing an equilibrium of forces, where $F_{net} = 0$.

Vector components of a translational equilibrium of forces

A vector diagram that results in a closed loop fulfils the conditions of a translational equilibrium of forces. If two forces are perpendicular, Pythagoras' theorem and trigonometry can be used to determine the magnitude and direction of a third force that will result in translational equilibrium with the two other forces.

If two forces are not perpendicular, each force is replaced with two vectors that are in perpendicular directions. These perpendicular vectors are called the **components** of a force. These force components are then added in each dimension, which results in two perpendicular resultant vectors that can be added using Pythagoras' theorem and trigonometry. The diagrams in Figure 11.2.4 show how two non-perpendicular vectors are added using vector diagrams.

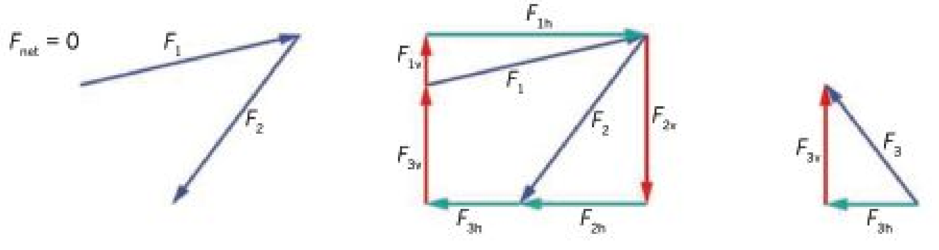


FIGURE 11.2.4 A method for finding F_3 that acts to make an equilibrium of forces with F_1 and F_2.

The guiding principle behind this method is that, for the forces to be in equilibrium, the sum of the *x*-axis (horizontal or left–right) forces must equal zero, and the sum of the *y*-axis (vertical or up–down) forces must also equal zero.

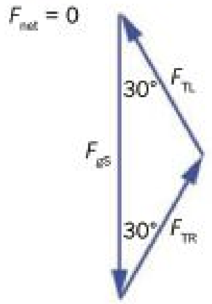
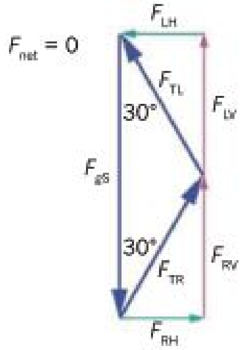
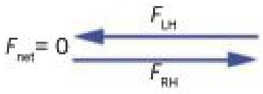
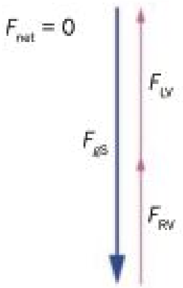
$$F_{net, x} = 0 \text{ or } F_{net, h} = 0 \text{ or } F_{net, left–right} = 0$$

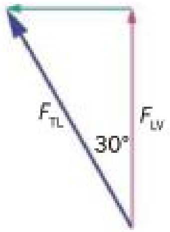
$$F_{net, y} = 0 \text{ or } F_{net, v} = 0 \text{ or } F_{net, up–down} = 0$$

Worked example 11.2.2

CALCULATING TRANSLATIONAL EQUILIBRIUM IN TWO DIMENSIONS

An advertising banner is hung by two cables to the ceiling of a shop. The banner has a mass of 45.0 kg and the cables are at an angle of 30° to the vertical as shown in the image below. If the mass of the cables is ignored, calculate the tension in each cable when the sign is suspended. Use $g = 9.80\ \text{N kg}^{-1}$ when answering this question.

Thinking	Working
Construct a vector diagram adding all of the forces together. F_{gS} is the force due to the weight of the sign. F_{TL} and F_{TR} are the tension forces in the wires on the left-hand side and right-hand side of the sign, respectively.	
Apply horizontal components and vertical components. F_{LH} and F_{LV} are the horizontal and vertical components of the tension force in the left-hand wire, and F_{RH} and F_{RV} are the horizontal and vertical components of the tension force in the right-hand wire.	
Recognise that, in the horizontal dimension, F_{LH} is in equilibrium with F_{RH}.	
Recognise that, in the vertical dimension, F_{gS} is in equilibrium with F_{RV} and F_{LV}.	

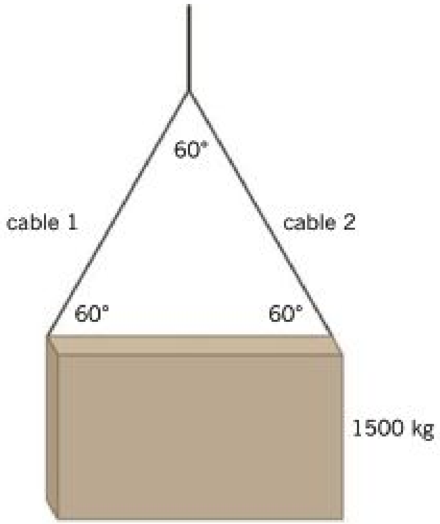
Apply the equation for translational equilibrium in the vertical dimension. Recognise that F_{RV} and F_{LV} are equal in magnitude and therefore each is half of F_{gS}.	$F_{net,V} = 0$ $F_{RV} = F_{LV}$
Expand the equation to include each of the vertical forces acting on the sign.	$F_{gS} + F_{RV} + F_{LV} = 0$ $m_S g + F_{RV} + F_{LV} = 0$
Substitute the data into the equation and solve for the unknown.	$45.0 \times -9.8 + F_{RV} + F_{LV} = 0$ $-441 + F_{RV} + F_{LV} = 0$ $F_{RV} + F_{LV} = 441$ $F_{RV} = F_{LV} = 220.5$ N
Draw the right triangle with one of the vertical components of the tension and the angle.	
Use trigonometry to solve for the tension in one of the cables, which will equal the tension in the other cable as well.	$\cos\theta = \frac{F_{LV}}{F_{TL}}$ $F_{TL} = \frac{F_{LV}}{\cos\theta}$ $= \frac{220.5}{\cos 30°}$ $= 255$ N $= F_{TR} = 255$ N

Worked example: Try yourself 11.2.2

CALCULATING TRANSLATIONAL EQUILIBRIUM IN TWO DIMENSIONS

A concrete beam of mass 1500 kg is being lifted by cables labelled 1 and 2, as shown in the diagram. The beam is moving upwards with a constant velocity of 2.0 m s^{-1}. Calculate the tension in cable 1 and cable 2. Ignore the mass of the cable and use $g = 9.80$ N kg^{-1} when answering this question.

Give your answers to three significant figures.



11.2 Review

SUMMARY

- A translational equilibrium of forces occurs when the sum of the forces acting on an object add to give a zero resultant force or zero net translational force.
- Translational equilibrium in one dimension can be represented mathematically as $\Sigma F = 0$ (or $F_{net} = 0$)
- Translational equilibrium in two dimensions can be represented mathematically as $F_{net,\,x} = 0$ and $F_{net,\,y} = 0$
- In two dimensions, an equilibrium of forces can be represented in a closed vector diagram.
- By calculating the *x* and *y* components of known forces, you can use the equilibrium equation in each dimension to find the unknown force required to keep an object in equilibrium.

KEY QUESTIONS

Use $g = 9.80$ N kg^{-1} when answering these questions.

1 Which one of the following is correct for an object to be in translational equilibrium?

A the object must be stationary
B the object must be moving at constant velocity
C the object must be experiencing translational acceleration
D the object must not be experiencing translational acceleration

2 Which one or more of the following are in translational equilibrium?

A a stationary elevator
B an elevator going up with constant velocity
C an aeroplane during take off
D a container ship sailing with constant velocity
E a car plummeting off a cliff

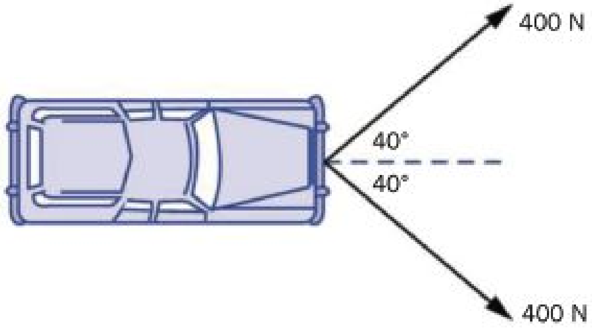
3 Calculate the resultant force in each of the following vector additions:

a 200 N up and 50 N down
b 65 N west and 25 N east
c 10 N north and 10 N south
d 10 N north and 10 N west

4 A single string supports a bird-feeder that has a mass of 355 g. If the bird-feeder is in translational equilibrium calculate the tension (upwards) force provided by the string to hold the bird-feeder.

5 Two window cleaners working on the windows of a high-rise building work on a platform that is suspended by four cables. Tom has a mass of 79 kg, Jack has a mass of 68 kg and the mass of the platform is 225 kg. If the platform is moving down the side of the building at a constant speed and each cable provides the same tension as the other cables, then calculate the tension in one of the cables.

6 A small car is pulled by two people using ropes. Each person supplies a force of 400 N at an angle of 40° to the direction in which the car travels.



a What is the net force applied by the people to the car?
b If friction is also acting on the car so that it is in translational equilibrium, determine the force due to friction.

7 A rectangular advertising sign is supported from its upper corners by two cables, each making an angle of 40° to the vertical. The sign has a mass of 5.0 kg. Calculate the tension in each cable.

11.3 Static equilibrium

Objects are in translational equilibrium when the sum of the forces acting on the object equals zero. However, some forces may act in ways that generate torque or moment on an object. This will depend on the point of application of the force(s). In order for an object to be in rotational equilibrium (see Figure 11.3.1), the sum of the moments in a clockwise direction must balance the sum of the moments in an anticlockwise direction. This relationship is called the **Principle of Moments**.

By combining the conditions for translational equilibrium and rotational equilibrium, an object can achieve a state of **static equilibrium**. Static equilibrium will be explored in more detail in this section.

FIGURE 11.3.1 When a cyclist doesn't need to pedal, they can stand on both pedals with equal force. This causes an equal torque in the clockwise and anticlockwise directions. The pedals are in rotational equilibrium.

CENTRE OF MASS

Think about an athlete running in a 100 m sprint. In simple terms, the athlete runs in a straight line along the track. The displacement and velocity of the athlete at any time can be calculated using the principles discussed in Chapter 9. In reality, however, the motion of the various parts of the athlete's body will differ significantly during the run. Her arms and legs move in a complex manner that is not easy to analyse.

The analysis of the motion of complicated systems such as a sprinter or high-jumper can be simplified to the motion of a single point. The mass of the sprinter can be considered to be 'concentrated' into a single point, which has travelled in a straight line. This single point is called the centre of mass. There is as much mass above the centre of mass as there is below it, as much mass to the left as there is to the right, and as much mass in front as there is behind it.

If an object is uniform in one dimension only (e.g. a straight piece of wire), its centre of mass will be exactly half way along its length. For an object that is uniform in two dimensions (e.g. a flat piece of cardboard), the centre of mass will be the point which is half way along and half way across the object. For an object that is uniform in three dimensions (e.g. a wooden block such as that shown in Figure 11.3.2), the centre of mass will be the point that is halfway along all dimensions of the object. It is possible for the centre of mass to lie outside the body. For example, the centre of mass of a doughnut is located at the centre of the hole. A person's centre of mass is typically just above their navel, but it will vary with the positions of the arms and legs.

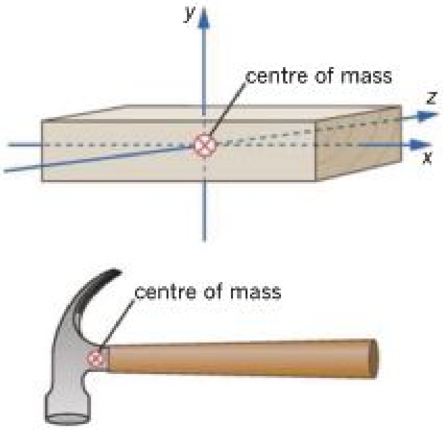


FIGURE 11.3.2 The centre of mass of a three-dimensional block of wood is shown with arrows drawn in the x, y and z dimensions illustrating the lines where there is equal mass on either side of the lines. The centre of mass of a non-uniform hammer shows that the centre of mass will be closer to the end with a higher density than the end with the lesser density.

A concept that is closely related to centre of mass is **centre of gravity**. Instead of being a point particle whose motion equates to the whole extended body or system, the centre of gravity is the position from which the entire weight of the body or system is considered to act. As a consequence of this, the centre of gravity is the position at which the body will balance. For all practical purposes, the centre of gravity is exactly at the centre of mass. It is only when a body is so large that it is in a non-uniform gravitational field that the centre of gravity no longer coincides with the centre of mass.

The centre of mass is an important concept when considering an object's stability.

STABILITY AND EQUILIBRIUM

Equilibrium refers to a state in which an object is balanced so that, as long as no external factors are changed, the object will not accelerate. Stability refers to the tendency of the system to maintain, or return to, the equilibrium state if external factors are changed. There are three types of equilibrium related to the stability of an object:

- **neutral equilibrium**—the object will remain stationary no matter where it is placed (e.g. placing a ball anywhere on a large flat surface)
- **stable equilibrium**—if the object is moved away from its equilibrium position, forces will act to return it to its equilibrium position (e.g. placing a ball in a large bowl—the equilibrium position is at the bottom of the bowl, and if it is moved to any other position, gravity will act to return it to the bottom of the bowl)
- **unstable equilibrium**—if the object is moved away from its equilibrium position, forces will act to push it further from the equilibrium position (e.g. placing a ball on top of a large dome—the equilibrium position is at the top of the dome, but if it is moved in any direction, gravity will pull it further from the top of the dome).

When designing structures, engineers and architects want to ensure that balance and stability are maintained. Whether this occurs depends on the relative positions of the centre of gravity and the **base** or point of support. When a vertical line downwards from the centre of gravity passes through the base of support, the object is stable. The vertical line from the centre of gravity represents the direction of the force of gravity on the object.

In Figure 11.3.3, the weight of the car passes through the car's support base, between the wheels. Taking the centre of gravity to be the pivot point, the reaction forces on the left wheels provide a clockwise torque while the reaction forces on the right wheels provide an anticlockwise torque. Therefore the torques will balance and the car will not tip over. In the case of the truck, however, the weight is directed outside the truck's base of support, so an anticlockwise torque acts to tip the truck over.

The stability of an object or structure can be increased in two ways:

- if the centre of gravity is lowered
- if the width of the support base is increased.

As a result of either of these, the object has to be tipped further to make the force of gravity act outside the support base. Racing cars have a very low centre of gravity to increase their stability when cornering at high speed. Training wheels on a child's bicycle widen the support base, making it harder to tip the bicycle sideways.

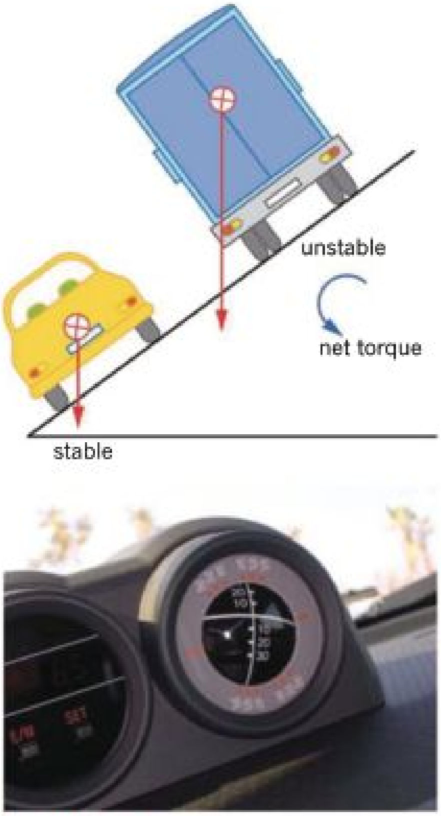


FIGURE 11.3.3 (a) The car on the incline is in stable equilibrium, while the heavily laden truck on the same incline could topple. The weight vector is outside the lower point of support for the truck, so there is no reaction force from the road to the higher wheel. (b) Modern four-wheel drives and tractors have inclinometers to warn the driver if the vehicle is in danger of tipping.

STATIC EQUILIBRIUM

Rotational equilibrium

A **rotational equilibrium** of torque or moments occurs about a **reference point** when the sum of all the torques acting on an object in the clockwise direction is equal to the sum of all the torques acting in the anticlockwise direction.

When analysing a situation where more than one torque acts on an object, the Principle of Moments applies where rotational equilibrium will exist if the net torque about a reference point is equal to zero.

$\tau_{\text{net}} = 0$

Recall that $\tau = r_{\perp}F$

where τ is the torque (N m)

$r_{\perp}$ is the force arm (m)

F is the force (N).

$\tau_{\text{net}} = 0$ implies that the sum of torques in the clockwise direction is balanced by the sum of torques in the anticlockwise direction. Therefore, the Principle of Moments is also represented as:

$$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$$

Rotational equilibrium is demonstrated in Figure 11.3.4.

When the sum of the torques is not balanced, then the object will rotate about the centre of mass.

FIGURE 11.3.4 To keep this mast in rotational equilibrium in relation to its base, stainless steel cables (stays) attached to it must provide opposing torques on the mast.

Conditions for static equilibrium

When a body or a system is not accelerating or rotating, it is in both translational and rotational equilibrium. This situation then satisfies the conditions for static equilibrium. This can be represented by:

$$F_{\text{net}} = 0 \text{ and } \tau_{\text{net}} = 0$$

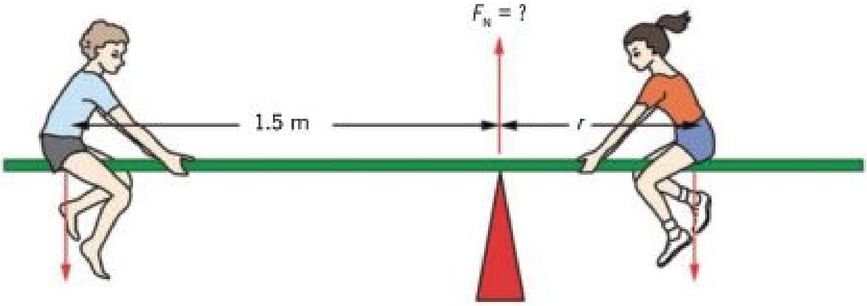
which can also be represented as:

$$F_{\text{net}} = 0 \text{ and } \Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$$

Worked example 11.3.1

CALCULATING STATIC EQUILIBRIUM

While playing in their backyard, two young children make a see-saw with a long plank. The boy sits on the see-saw 1.50 m from the pivot. The masses of the boy and girl are 20.0 kg and 30.0 kg, respectively. Assume that the plank's mass is negligible. Use $g = 9.80\ \text{N kg}^{-1}$ when answering these questions.



a Calculate the force applied to the plank due to the pivot point when the children are sitting on the see-saw.

Thinking	Working
Identify the variables involved and state them with their directions in their standard form.	$m_g = 30.0$ kg $m_b = 20.0$ kg $g = 9.80\ \text{N kg}^{-1}$ down
Apply a sign convention to the vector data.	$g = -9.80\ \text{N kg}^{-1}$

Identify the object that is in translational equilibrium. This is the object on which all the forces are acting.	The object experiencing translational equilibrium is the see-saw.
Apply the equation for translational equilibrium in the vertical dimension.	$F_{\text{net}, y} = 0$
Expand the equation to include each of the forces acting on the plank.	$F_g + F_b + F_P = 0$ $m_g g + m_b g + F_P = 0$
Substitute the data into the equation and solve for the unknown.	$(30.0 \times -9.80) + (20.0 \times -9.80) + F_P = 0$ $(-294) + (-196) + F_P = 0$ $-490 + \text{Fp} = 0$ $F_P = 490$ N
State the answer with the appropriate direction.	$F_P = 490$ N up

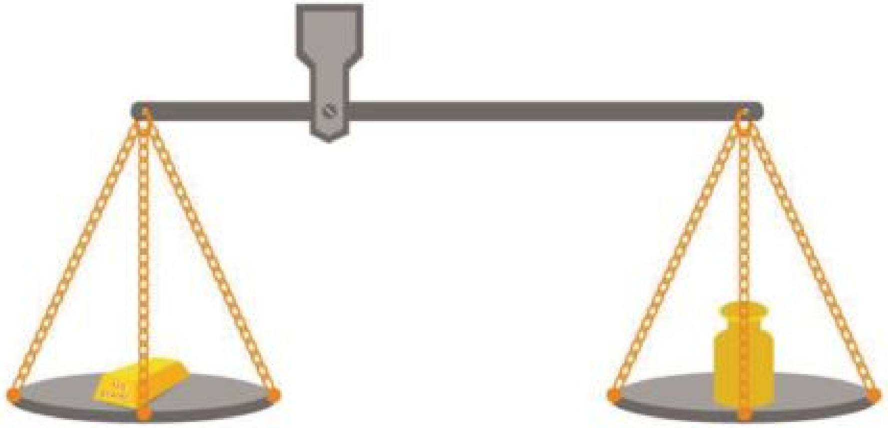
b Calculate where the girl has to sit in order to balance the boy.	
Thinking	**Working**
Identify the variables involved and state them in their standard form.	$m_g = 30.0$ kg $m_b = 20.0$ kg $r_{\perp b} = 1.50$ m $g = 9.80$ N kg^{-1}
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the see-saw.
Decide the reference point about which the torques will be calculated.	The reference point is the pivot of the see-saw.
Decide which force causes the clockwise torque and which force causes the anticlockwise torque around the chosen reference point.	The force of the girl on the see-saw provides the clockwise torque. The force of the boy on the see-saw provides the anticlockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the see-saw.	$r_{\perp g} F_g = r_{\perp b} F_b$
Substitute the data into the equation and solve for the unknown.	$F_g r_{\perp g} = F_b r_{\perp b}$ $r_{\perp g} m_g g = r_{\perp b} m_b g$ $r_{\perp g} 30.0 \times 9.80 = 1.50 \times 20.0 \times 9.80$ $r_{\perp g} = \dfrac{20.0 \times 9.80 \times 1.50}{30.0 \times 9.80}$ $= 1.00$ m

Worked example: Try yourself 11.3.1

CALCULATING STATIC EQUILIBRIUM

A set of scales (with one longer arm) is used to measure the mass of gold. A lump of gold with a mass of 150 g is placed on the short-arm, which is 10.0 cm long, and a standard set of masses are placed on the long-arm.

Use $g = 9.80\ \text{N kg}^{-1}$ when answering these questions. Give your answers to three significant figures.



a Calculate the force applied to the scale's arm due to the pivot point if a standard mass of 50.0 g exactly balances the gold.

b Calculate length the long arm should have in order to balance the gold.

In Worked example 11.3.1, the see-saw is in equilibrium because all the forces and torques are balanced. In solving the problem, it seemed obvious to choose the pivot as the reference point, around which the torques are determined. But because the see-saw plank is in equilibrium, any point could have been chosen as the reference point. For example, the reference point could be where the girl is sitting (X) as shown in Figure 11.3.5. This will mean that the torques acting on the plank would be due to the boy, and due to the see-saw pivot point. The torque due to the girl becomes zero as the lever arm distance for her will be zero.

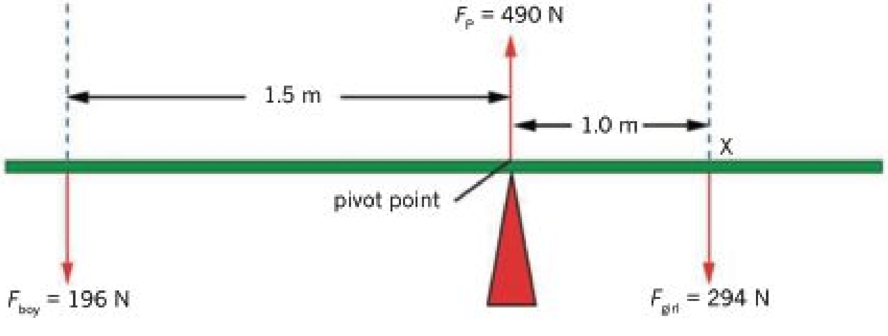


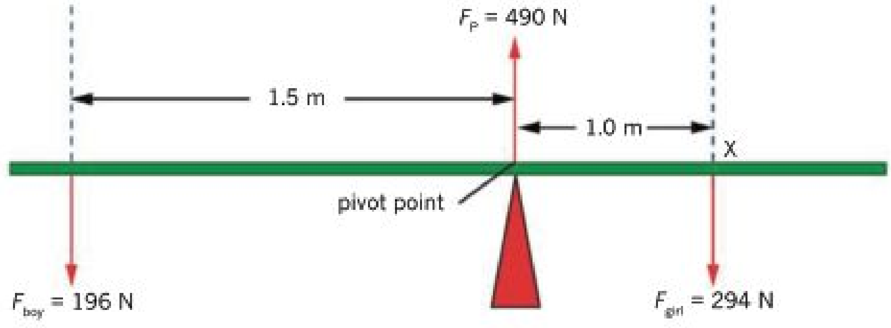
FIGURE 11.3.5 A force diagram for the see-saw problem from Worked example 11.3.1 above, where the point at which the girl sits (labelled X) has been chosen as the reference point.

The boy will create an anticlockwise torque around the girl, and the normal force at the pivot for the see-saw, F_P, creates an equal clockwise torque around the girl. In Worked example 11.3.2 the see-saw is in rotational equilibrium. This is verified by calculating the torques around the position of the girl. The plank will be in rotational equilibrium if the clockwise torque equals the anticlockwise torque.

Worked example 11.3.2

CALCULATING STATIC EQUILIBRIUM USING A DIFFERENT REFERENCE POINT

Verify that the see-saw plank in the image below is in rotational equilibrium about the reference point X, where the girl is sitting. The weight of the boy and girl are 196 N and 294 N, respectively, and the force of the pivot on the plank is 490 N upwards. Assume that the plank's mass is negligible.

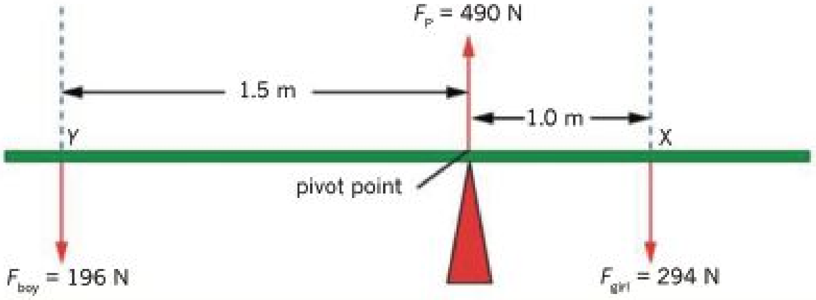


Thinking	**Working**
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the see-saw.
Identify the variables involved and state them in their standard form.	$F_P = 490$ N $F_g = 294$ N $F_b = 196$ N $r_{\perp b} = 2.50$ m $= 1.00$ m
Decide the reference point about which the torques will be calculated.	The reference point is the position of the girl, at X.
Decide which force causes the clockwise torque, and which force causes the anticlockwise torque around the chosen reference point.	The force of the pivot on the plank provides the clockwise torque. The force of the boy on the plank provides the anticlockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the see-saw. Note that the girl's torque is not included here as her torque is zero.	$r_{\perp P}F_P = r_{\perp b}F_b$
Substitute the data into the equation and evaluate.	$r_{\perp P}FP = r_{\perp b}F_b$ $1.00 \times 490 = 2.5 \times 196$ $490 = 490$
Describe the magnitude of the clockwise torque compared to the magnitude of the anticlockwise torque.	Around reference point X (the position of the girl), the clockwise torque due to the pivot point on the plank is equal to the anticlockwise torque due to the boy on the plank. So the plank is in rotational equilibrium.

Worked example: Try yourself 11.3.2

CALCULATING STATIC EQUILIBRIUM USING A DIFFERENT REFERENCE POINT

Verify that the see-saw plank in the figure below is also in rotational equilibrium about the reference point where the boy is sitting, at point Y. The weights of the boy and girl are 196 N and 294 N, respectively, and the force of the pivot on the plank is 490 N upwards. Assume that the plank's mass is negligible.



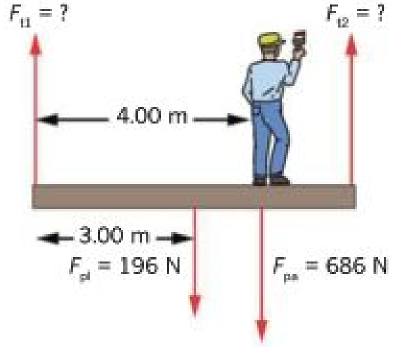
Static equilibrium with two unknown forces

The see-saw problem is relatively straightforward since, in each situation, there is only one unknown force. If there are two unknown forces, the reference point can be chosen to coincide with one of the forces. By using this strategy it means that the force acting at the reference point contributes no torque (since $r = 0$). The remaining unknown force can be found using the relationship $\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$. Worked example 11.3.3 below employs this strategy.

Worked example 11.3.3

CALCULATING STATIC EQUILIBRIUM WITH TWO UNKNOWNS

While painting a tall building, a 70.0 kg painter stands 4.00 m from the end of a 6.00 m–long plank that is supported by a rope at either end. The plank has a mass of 20.0 kg. Use $g = 9.80\ \text{N kg}^{-1}$ when answering this question.



Determine the tension on the left-hand rope (F_{t1}).

Thinking	Working
Decide the reference point about which the torques will be calculated. Note that this must be the point at which the other unknown force acts.	The reference point is the point at which the rope providing the tension force, F_{t2}, is attached.
Identify the variables involved and state them in their standard form. Remember to quote all $r_{\perp}$ lengths from the reference point, F_{t2}. Take the centre of the plank to be the point at which its weight acts.	$m_{pl} = 20.0$ kg $m_{pa} = 70.0$ kg $r_{\perp Ft1} = 6.00$ m $r_{\perp pl} = 3.00$ m $r_{\perp pa} = 2.00$ m $g = 9.80\ \text{N kg}^{-1}$

Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the plank.
Decide which force causes the clockwise torques and which forces cause the anticlockwise torques around the chosen reference point.	The tension force of the left hand rope on the plank provides the clockwise torque. The force of the painter on the plank provides an anticlockwise torque. The force of gravity on the plank provides another anticlockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the plank. Note the torque of the right-hand rope is not included as it acts through the reference point.	$r_{\perp Ft1}F_{t1} = r_{\perp pl}F_{pl} + r_{\perp pa}F_{pa}$
Substitute the data into the equation and solve for the unknown.	$6.00 \times F_{t1} = 3.00 \times 20 \times 9.80 + 2.00 \times 70 \times 9.80$ $F_{t1} = \frac{(3.00 \times 20 \times 9.80) + (2.00 \times 70 \times 9.80)}{6.00}$ $= \frac{588 + 1372}{6.00}$ $= 327\text{ N}$

Worked example: Try yourself 11.3.3

CALCULATING STATIC EQUILIBRIUM WITH TWO UNKNOWNS

For the painter on the plank scenario above, determine the tension on the right hand rope (F_{t2}).

Another way to determine the second unknown force is to apply the conditions for translational equilibrium. You can check these values using $F_{\text{net}} = 0$: the sum of the two upwards forces (tensions), 555 N + 327 N = 882 N. This balances the sum of the two downwards forces, 196 + 686 = 882 N.

OTHER STATIC EQUILIBRIUM SCENARIOS

The see-saw scenario and the supported plank are just two of many situations in which static equilibrium can be used to solve the forces involved. The conditions of static equilibrium will be applied to the following scenarios.

Cantilevers

A beam that extends beyond its support structure is called a **cantilever**. Cantilevers are common structural elements. For example, the diving board at the local pool is a cantilever. A cantilever bridge might be used to span a river or valley, as shown in Figure 11.3.6. Pillars are built at regular intervals across a river in order to support a number of beams. The cantilever beams are then joined at the centre of each span. The forces on the pillars are not affected by joining the beams; these are the same as if the beams were not connected. All the support for the cantilever is supplied by pillars. Other structures that can involve cantilevers include shelving, awnings over the footpath outside some shops and the wings of a plane.

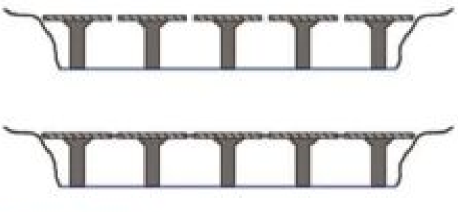


FIGURE 11.3.6 Each beam in the cantilever bridge is fully supported by the pillar below the beam. By connecting the beams no added support is provided.

When the centre of mass of the beam is not directly above a support, then the force due to gravity that acts on the centre of mass will provide a torque. This must be factored into the conditions of rotational equilibrium. In this case two supports are usually required, with one support providing an upwards force on the beam and the other support providing a downwards force on the beam.

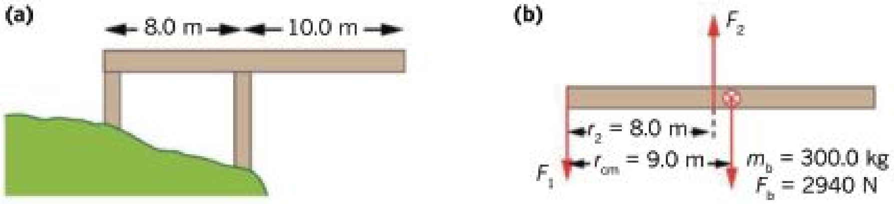
FIGURE 11.3.7 A diving board showing that a metal strap is needed to provide the downwards force on the fixed end of the board.

Figure 11.3.7 shows a swimming-pool diving board, with one support providing an upwards force on the board and the other support providing a downwards force on the board. A diving board must have a downwards force on the fixed end of the board to provide an opposing torque to the torque provided by the force due to gravity on the board.

Worked example 11.3.4

USING STATIC EQUILIBRIUM TO CALCULATE THE FORCE ON A CANTILEVER

A uniform cantilever beam 18.0 m long is used as a viewing platform. It extends 10.0 m beyond two supports which are 8.00 m apart. The beam has a mass of 300.0 kg.



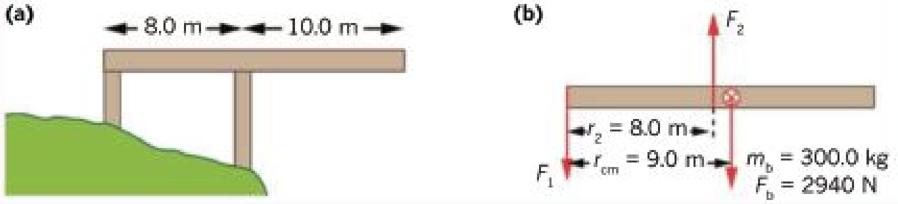
Determine the magnitude and direction of the force that the right-hand support must supply so that the beam is in static equilibrium (F_2).

Thinking	Working
Decide the reference point about which the torques will be calculated. Note that this must be the point at which the other unknown forces act.	The reference point is the point at which the left-hand support providing the force F_1 is attached.
Identify the variables involved and state them in their standard form.	$m_b = 300.0$ kg $r_{\perp F2} = 8.00$ m $r_{\perp cm} = 9.00$ m $g = 9.80$ N kg^{-1}
Identify the object that is in rotational equilibrium. This is the object upon which all the torques are acting.	The object experiencing rotational equilibrium is the beam.
Decide which force causes the clockwise torque and which force causes the anticlockwise torque around the chosen reference point.	The force of the right-hand support on the beam provides the anticlockwise torque. The force of gravity on the beam provides the clockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the beam. Note the torque of the left-hand support is not included, as it acts through the reference point.	$F_b r_{\perp cm} = F_2 r_{\perp F2}$
Substitute the data into the equation and solve for the unknown.	$300 \times 9.80 \times 9.00 = F_2 \times 8.00$ $F_2 = \frac{300 \times 9.80 \times 9.00}{8.00}$ $= \frac{26460}{8.00}$ $= 3308$ N
State the direction of the force acting on the object in equilibrium.	force acts upwards on the beam.

Worked example: Try yourself 11.3.4

USING STATIC EQUILIBRIUM TO CALCULATE THE FORCE ON A CANTILEVER

Determine the magnitude and direction of the force that the left-hand support must supply so that the beam is in static equilibrium (F_1).



Another way to determine the second unknown force is to apply the conditions for translational equilibrium. We can check these values using $\Sigma F = 0$: the sum of the upwards force (F_2) is 3308 N. This balances the sum of the two downwards forces (F_b and F_1): 2940 + 368 = 3308 N.

Struts and ties

As well as their main beams and pillars, many structures have additional ways to strengthen them. A structure, such as a cantilever, may be supported by struts and ties. A strut will be under compression (squeezed) and must be rigid. A tie will be under tension (stretched) and may be rigid or flexible. Struts and ties are shown in Figure 11.3.8.

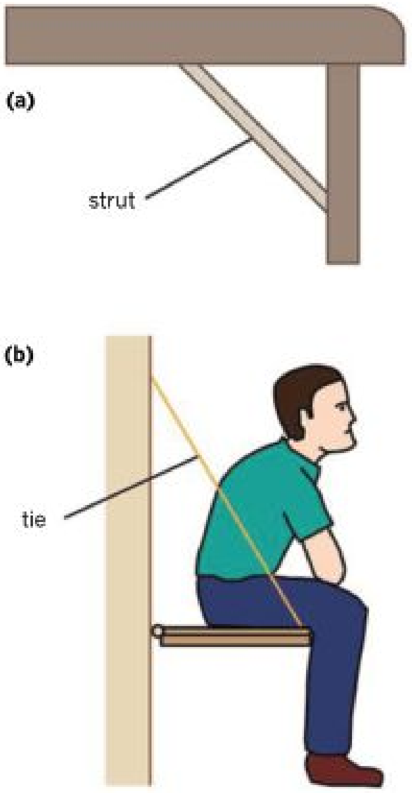
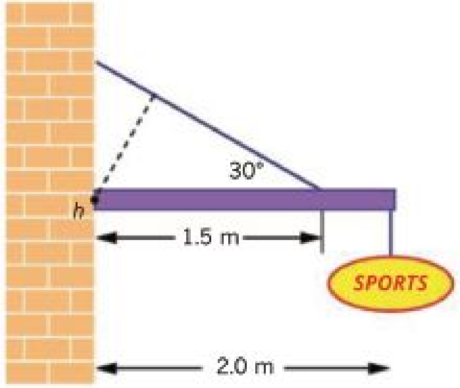


FIGURE 11.3.8 A strut (a) helps to support a cantilevered beam and is under compression. (b) A tie helps to support a fold-out bench and is under tension.

Worked example 11.3.5

USING STATIC EQUILIBRIUM TO CALCULATE THE TENSION IN A TIE THAT IS SUPPORTING A BEAM

A sign of mass 10 kg is suspended from the end of a uniform 2.0 m long cantilevered beam. The other end of the beam is attached to the wall by a hinge labelled *h*. The beam has a mass of 25 kg and is further supported by a wire tie that makes an angle of 30° to the beam. The wire is attached to the beam at a point 1.5 m from the wall. Use $g = 9.80 \text{ N kg}^{-1}$ and ignore the mass of the wire for these calculations.



Calculate the tension (F_t) in the wire that is supporting the beam.

Thinking	Working
Decide the reference point about which the torques will be calculated. Note that this must be the point at which the other unknown force acts.	The reference point is the hinge (*h*) where the beam is connected to the wall.

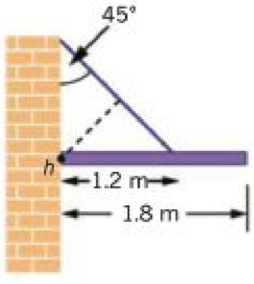
Identify the variables involved and state them in their standard form. Here the subscripts used are: b = beam, c = centre of mass of the beam, s = sign and w = wire.	$m_b = 25.0$ kg $m_s = 10.0$ kg $r_{\perp c} = 1.0$ m $r_{\perp s} = 2.00$ m $r_w = 1.50$ m $g = 9.80\text{ N kg}^{-1}$
Identify the object that is in rotational equilibrium. This is the object on which all the torques are acting.	The object experiencing rotational equilibrium is the beam.
Decide which force causes the anticlockwise torque and which forces cause the clockwise torque around the chosen reference point.	The force of the wire tie on the beam provides the anticlockwise torque. The force of gravity on the beam provides one clockwise torque. The force of gravity on the sign provides another clockwise torque.
Apply the equation for rotational equilibrium.	$\Sigma\tau_{\text{clockwise}} = \Sigma\tau_{\text{anticlockwise}}$
Expand the equation to include each of the torques acting on the beam.	$r_{\perp c}F_b + r_{\perp s}F_s = r_{\perp w}F_w$
Solve for the perpendicular distances from the force arm to the line of action of the force.	$r_{\perp w} = r_w \sin 30°$ $= 1.5 \times \sin 30°$ $= 0.75$ m
Substitute the data into the equation and solve for the unknown force.	$(1.00 \times 25 \times 9.80) + (2.00 \times 10 \times 9.80) = F_t \times 0.75$ $F_t = \dfrac{245 + 196}{0.75}$ $= \dfrac{441}{0.75}$ $= 588$ N Note that by finding the perpendicular distance the tension force can be calculated directly by the equation. If the vertical component of the tension force was found then the tension force would need to be calculated.

Worked example: Try yourself 11.3.5

USING STATIC EQUILIBRIUM TO CALCULATE THE TENSION IN A TIE THAT IS SUPPORTING A BEAM

A uniform 5.00 kg beam, 1.80 m long, extends from the side of a building and is supported by a wire tie which is attached to the beam 1.20 m from a hinge (*h*) at an angle of 45°. Determine the tension in the cable.

Give your answers to three significant figures.



Calculate the tension (F_t) in the wire that is supporting the beam.

11.3 Review

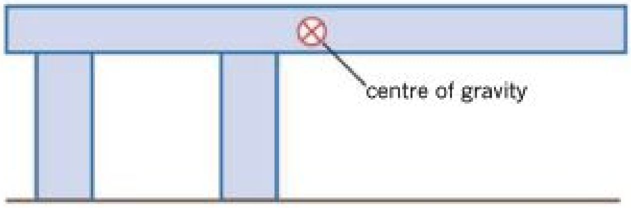
SUMMARY

- The centre of mass is a single point in an object where the mass can be considered to be 'concentrated' for the purposes of analysing motion.
- Static equilibrium occurs when an object experiences translational equilibrium and rotational equilibrium.
- Static equilibrium can be represented mathematically as $F_{net} = 0$ (translational equilibrium) and $\tau_{net} = 0$ (rotational equilibrium).
- Rotational equilibrium can be represented mathematically as $\Sigma\tau_{clockwise} = \Sigma\tau_{anticlockwise}$
- In calculations of static equilibrium with only one unknown force the reference point is the point about which the torques act.
- In calculations of static equilibrium with two unknown forces, the reference point can be placed at the point at which one of the unknown forces acts. This eliminates any torque due to this force as the distance from the force to the reference point is zero.

KEY QUESTIONS

Use $g = 9.80\ \text{N kg}^{-1}$ when answering these questions.

1 A student designs a bench that consists of a long beam resting on top of two supports, as shown in the diagram. The centre of gravity of the beam is indicated. The student does not use any nails, screws, bolts, ropes or adhesives. Explain whether the bench will work successfully and what could be done, if anything, to improve the design.



2 Which one of the following describes the torques acting on an object in rotational equilibrium?

A It must have only clockwise torques.
B It must have only anticlockwise torques.
C It must have unequal clockwise and anticlockwise torques.
D It must have equal clockwise and anticlockwise torques.

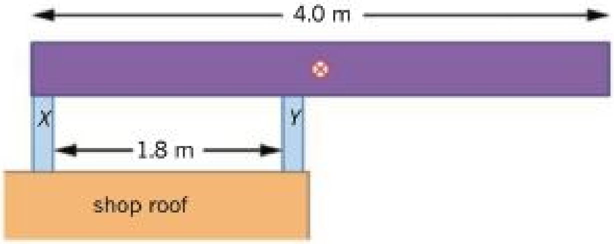
3 A cyclist is coasting down a road at a constant velocity while standing on the pedals. Which object is in static equilibrium in the list below?

A the rear wheel
B the front wheel
C the front cog connected to the pedals
D the front and rear wheels

4 An adult of mass 75.0 kg sits on a see-saw at a playground with a 25.0 kg child sitting 2.25 m from the pivot point on the other side from the adult. Calculate where the adult must sit for the see-saw to remain balanced and horizontal.

5 A makeshift shelf is used in a bakery to store sacks of flour. The shelf is constructed using a 10.0 m beam with a mass of 50.0 kg, with a support positioned at each end. The shelf holds sacks of mass 100 kg, 150 kg and 200 kg at the positions shown in the figure below. Calculate the forces on the beam due to the left and right supports.

6 A 4.0 m cantilever-type awning is constructed on the roof of a shop so that it shades the front. The awning has a mass of 900 kg and is supported by two columns, X and Y, which produce forces on the awning of F_X and F_Y, respectively.

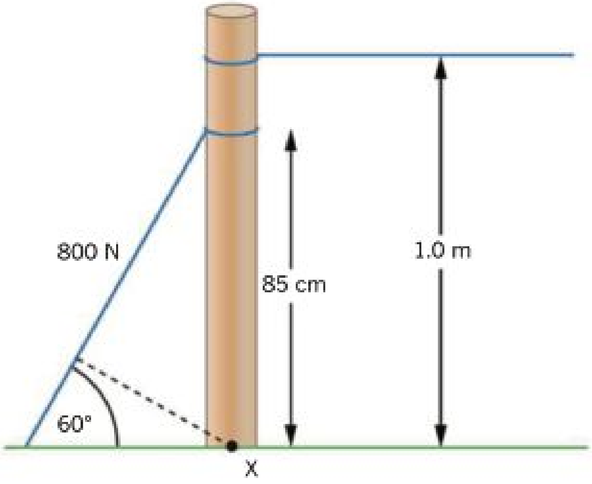


a Determine the direction in which F_X acts on the awning.
b Determine the force supplied by column Y on the awning (F_Y).
c Determine the force supplied by column X on the awning (F_X).

11.3 Review *continued*

7 An adult of mass 90.0 kg sits on a see-saw at a playground with two 20.0 kg children. The adult sits on the opposite side from the children. One child is sitting 1.50 m from the pivot point and the other sitting 2.50 m from the pivot point. Calculate where the adult must sit for the see-saw to remain balanced and horizontal.

8 The end-post of a fence is held in position by a backstay wire that is under a tension of 800 N at an angle of 60° to the horizontal. A horizontal fence wire connects to the other posts in the fence. A diagram of the fence is below.



a Determine the values for the horizontal and vertical components of the tension in the backstay wire.

b By considering the base of the post to be a pivot point, determine the size of the tension in the fence wire, F_t.

Chapter review

11

KEY TERMS

axis of rotation
base
cantilever
centre of gravity
components
force arm
line of action of the force
neutral equilibrium
pivot point
Principle of Moments
reference point
rotational equilibrium
stable equilibrium
static equilibrium
torque
translational equilibrium
unstable equilibrium

Use $g = 9.80\ \text{N kg}^{-1}$ when answering these questions.

1 Which option below explains the factors affecting the torque on an object?

A Torque only increases if the force applied to the object is increased.

B Torque increases when either the force or the force arm increases, or both increase.

C Torque only increases when the force arm decreases.

D Torque only increases when the force arm increases.

2 Calculate the torque applied to a bolt by a 20.0 cm–long spanner that has a force of 25.0 N acting at 90° to its length and at the end of the spanner.

3 Calculate the radius of the wheel on a pressure valve that supplies a torque on the valve of 3.47 N m when a force of 12.0 N is applied.

4 Which of the following options would provide you with the greatest torque when opening a 1 metre–wide door?

A pushing with a force of 33 N at right angles to the door, in the middle of the door

B pushing with a force of 25 N at right angles to the door, 25 cm from the handle edge of the door

C pushing with a force of 50 N at right angles to the door 25 cm from the hinges of the door

D pushing with a force of 25 N at right angles to the door, at the handle edge of the door

5 Demolishers wish to knock over a concrete wall. They plan to use a wrecking-ball which exerts a 5 kN force as it hits the wall. If it hits at a point that is 3 m above the ground, calculate the torque that is developed on the wall if it pivots at its base.

6 Calculate the length of a spanner that is used to tighten a nut to a torque of 32.1 N m when a force of 24.0 N is applied at right angles to the spanner, at the end of the spanner.

7 Which of the following is the correct description of the maximum torque on a door?

A The maximum torque will be achieved when the force is at a 45° angle to the door.

B The maximum torque will be achieved when the force is parallel to the door.

C The maximum torque will be achieved when the force is at a 90° angle to the door.

D The maximum torque will be achieved when the force is at a 100° angle to the door.

8 A camper ties a rope from the top of a 2.00 m tent pole to a peg on the ground. The rope is tightened so that the rope applies a 30.0 N force at an angle of 40.0° to the pole. Calculate the torque that is applied on the tent pole due to the rope if it pivots at its base.

9 Calculate the torque on a 45.0 cm spanner that is used to tighten a nut when a force of 62.0 N is applied at 65.0° to the spanner, at the end of the spanner.

10 Which of the following descriptions of calculating torque is correct?

A Torque can be calculated using either the perpendicular force or the perpendicular force arm.

B Torque can only be calculated using the perpendicular force.

C Torque can only be calculated using the perpendicular force arm.

D Torque can only be calculated using both the perpendicular force and the perpendicular force arm together.

11 A rope is attached at 30.0° to a freshly planted tree. The line of action of the force is along the same line as the rope, and the rope is attached to the tree 1.50 m up from the bottom of the tree. Assume the base of the tree is the pivot point.

a Calculate the length of the perpendicular force arm of the rope.

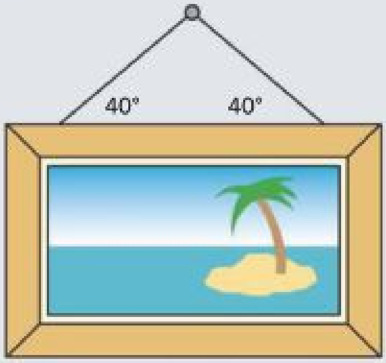
b Calculate the torque on the tree if the force applied to the tree by the rope is 12.5 N.

Chapter review *continued*

12 A young couple are assembling some flat-pack furniture with an Allen key. Calculate the torque supplied on a screw by a 7.00 cm long Allen key that has a force of 8.50 N acting at 90° to its length and at the end of the Allen key.

13 A mechanic uses a trolley jack to jack-up a car. The mechanic pushes vertically down on the end of a 90.0 cm lever with a force of 82.0 N as shown in the diagram below. The lever is shown at an angle of 40.0° up from the horizontal. Calculate the torque acting on the pivot.

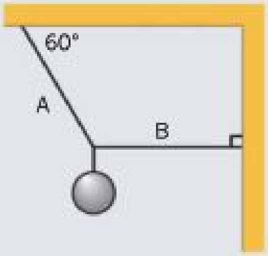
14 A picture is hung on a wall as shown in the diagram below. If the hanging wire is capable of supporting a maximum force of 40 N, what is the maximum mass of the picture that can be supported before the wire snaps?



15 A street performer stands on a rope tied between two bollards. When the performer is standing at the centre, the rope makes an angle of 10.0° to the horizontal. Assuming the mass of the rope is negligible and that the mass of the performer is 75.0 kg, calculate the tension in the rope.

16 Calculate the mass of a pendulum bob that hangs stationary from a chain if the tension in the chain is 7.50 N.

17 A bowling alley manager wants to promote their business by suspending a 100 kg fibreglass bowling ball on a frame outside their alley. The bowling ball is supported by two steel cables capable of withstanding up to 1500 N of tension before breaking. Cable A is at an angle of 60° to the horizontal frame member and cable B is perpendicular to the vertical frame member. Assuming the mass of the cables is negligible calculate the tension in cable A and in cable B.



18 Which of the following correctly describes rotational equilibrium?

- **A** A net torque acts about the reference point and rotation does not occur.
- **B** A net torque acts about the reference point and rotation occurs.
- **C** No net torque acts about the reference point and rotation does not occur.
- **D** No net torque acts about the reference point and rotation occurs.

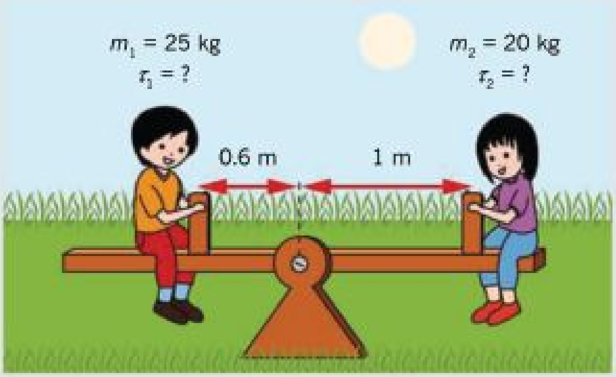
19 When an object is in static equilibrium, it experiences:

- **A** rotational equilibrium, but not translational equilibrium
- **B** rotational equilibrium and translational equilibrium
- **C** neither rotational equilibrium nor translational equilibrium
- **D** translational equilibrium, but not rotational equilibrium

20 Tom is riding his skateboard and is in a state of translational equilibrium. Which one or more of the following statements could be true of Tom's motion?

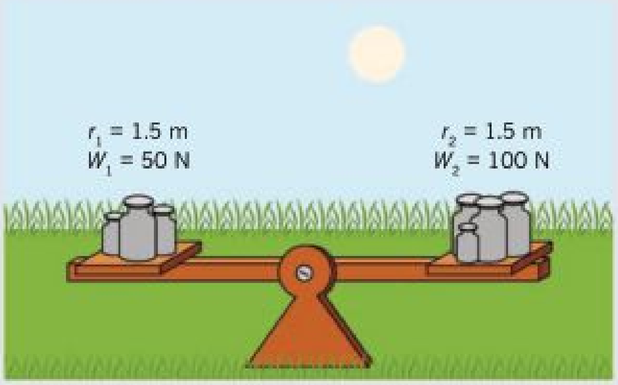
- **A** He is maintaining a constant velocity.
- **B** He is stationary.
- **C** He is experiencing a translational acceleration.
- **D** He is experiencing a rotational acceleration.

21 For the situation shown below:



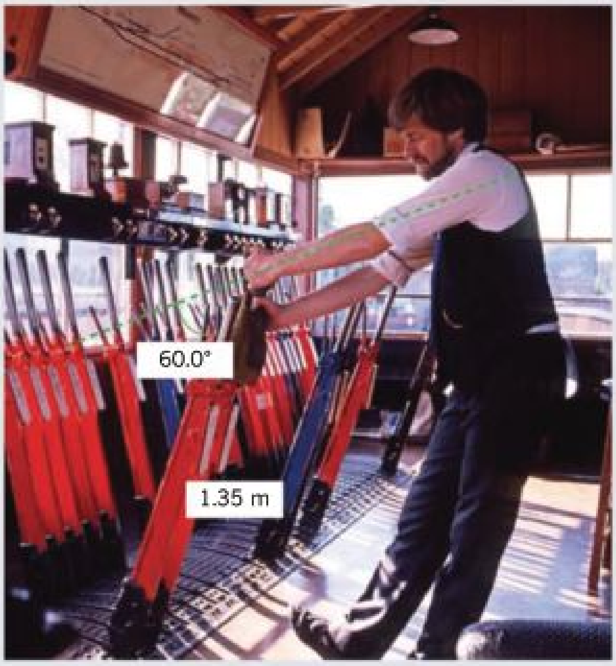
- **a** Give the magnitude and direction of the torque, τ_1, on the see-saw caused by the weight of the boy on the left.
- **b** Give the magnitude and direction of the torque, τ_2, on the see-saw caused by the weight of the girl on the right.
- **c** Give the magnitude and direction of the net torque acting on the see-saw.

22 For the situation shown below, determine the magnitude and direction of the resultant torque acting on the see-saw.



23 A child pulls down on a lever-type door handle with a force of 30.0 N. If the length of the handle is 12.0 cm then calculate the torque acting on the handle's pivot. Assume the force is applied at right angles to the handle.

24 A signalman is responsible for pushing levers that move train tracks to switch a train from one line to another, as shown in the figure below. Calculate the torque on the axle at the bottom of a 1.35 m lever if a 64.0 N force acts at an angle of 60.0° to the lever.



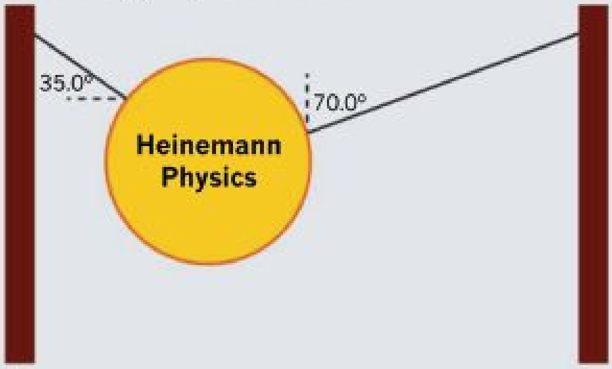
Chapter review *continued*

25 Hand-pumps are capable of pumping water from wells instead of using a bucket and rope. A man pushes horizontally on the end of a 105 cm lever with a force of 50.0 N, as shown in the diagram below. The lever is shown at an angle of 78.0° down from the horizontal. Calculate the torque acting on the pivot.

26 A climber is hanging from a rope half-way up a rock wall. If the climber has a mass of 86.5 kg, calculate the tension force provided by the rope in order for the climber to be in translational equilibrium.

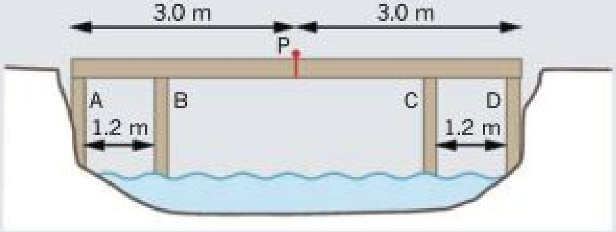
27 A dining table top of mass 40.0 kg is supported by four legs. On the table there is 4.50 kg of food. If the table is stationary, and each leg provides the same support as each of the other legs, then calculate the force of one of the legs on the table.

28 A 12.5 kg sign is hung between two poles by two cables of negligible mass, as shown in the diagram below. The sign is in translational equilibrium. Calculate the tension in the shorter left-hand cable and the longer right-hand cable.



29 A child's toy is suspended above her bed from a string attached to the ceiling. The toy is made from a 1.00 m aluminium rod of mass 10.0 g, with a 145 g Sun hanging at one end of the rod and a 22.5 g Earth hanging at the other end. Calculate how far from the Sun, in centimetres, the string needs to be tied to the bar so that the whole toy is in static equilibrium.

30 A pedestrian bridge over a small creek is made from two identical 3.0 m cantilevers, each of mass 400 kg.



a Calculate the reaction forces produced by the pillars A, B, C and D when:

- **i** there are no pedestrians on the bridge
- **ii** a 70 kg woman stands at position P with half her weight on each cantilever.

b What happens to the values of the forces in A and B as the woman walks from A past B to P?

CHAPTER

12

Energy, work and power

Throughout this chapter you will be learn about the common thread of energy conversion that is present in so many daily activities, as well as some more extreme activities. Your own personal energy stores are burnt up climbing steps or running to catch a bus. In more thrill-seeking adventures, such as bungee jumping, gravitational potential energy is converted into kinetic and elastic potential energy. Even jumping from a plane, the laws of Physics cannot be switched off!

At the end of this chapter, you will be able to define and use the terms *work*, *energy* and *power*. You will use force–displacement graphs to determine the amount of work done.

Key knowledge

By the end of this chapter, you will have covered material from the study of energy, work and power, and will be able to:

- apply the concept of work done by a constant force using:
  - work done = constant force × distance moved in direction of force: $W = Fs$
  - work done = area under force–distance graph
- investigate and analyse theoretically and practically Hooke's Law for an ideal spring: $F = -k\Delta x$
- analyse and model mechanical energy transfers and transformations using energy conservation:
  - changes in gravitational potential energy near Earth's surface: $E_g = mg\Delta h$
  - potential energy in ideal springs: $E_s = \frac{1}{2}k\Delta x^2$
  - kinetic energy: $E_k = \frac{1}{2}mv^2$
- analyse rate of energy transfer using power: $P = \frac{E}{t}$
- calculate the efficiency of an energy transfer system: $\eta = \frac{\text{useful energy out}}{\text{total energy in}}$

12.1 Work

The words 'energy' and 'work' are commonly used to describe a variety of everyday situations. However, these words take on quite specific definitions when used in a scientific context. They are two of the most important concepts in physics, allowing physicists to explain phenomena on a range of scales from collisions of subatomic particles to the interactions of galaxies.

ENERGY

Energy is the capacity to cause change. A moving car has the capacity to cause a change if it collides with something else. Similarly, a heavy weight lifted by a crane has the capacity to cause a change if it is dropped. Energy is a scalar quantity; it has magnitude but not direction.

There are many different forms of energy. **Mechanical energy** is defined as the energy that a body possesses due to its position or motion. This category of energy can be broadly classified into two groups: kinetic energy and potential energy.

Kinetic energy is energy associated with motion. Any moving object, like the moving car in Figure 12.1.1, has kinetic energy. In some forms of kinetic energy, the moving objects are not easily visible. An example of this is thermal energy, which is a type of kinetic energy related to the movement of particles. Table 12.1.1 lists some different types of kinetic energy and their associated moving objects.

FIGURE 12.1.1 A moving car has kinetic energy.

Type of kinetic energy	Moving objects
translational kinetic	objects moving in a straight line
rotational kinetic	rotating objects
thermal	atoms, ions or molecules
sound	air molecules

TABLE 12.1.1 Types of kinetic energy and their associated moving objects.

Potential energy is energy associated with the position of objects relative to one another or within fields. For example, an object suspended by a crane has **gravitational potential energy** because of its position in the Earth's gravitational field. Some examples of potential energy are listed in Table 12.1.2.

Type of potential energy	Cause
gravitational	gravitational fields
chemical	relative positions of atoms
magnetic	magnetic fields
nuclear	forces within the nucleus of an atom
elastic	attractive forces between atoms

TABLE 12.1.2 Types of potential energy and their causes.

Unit of energy

The SI unit for energy, the joule (J), is named after the English scientist James Prescott Joule. He was the first person to show that kinetic energy could be converted into heat energy. The energy represented by 1 J is the equivalent to the energy needed to lift a 1 kg mass (e.g. 1 L of milk) through a height of 0.1 m or 10 cm. More commonly, scientists work in units of kilojoules (1 kJ = 1000 J) or even megajoules (1 MJ = 1 000 000 J).

WORK

Although in everyday life the word 'work' can take on a variety of meanings, in a scientific context work has a very specific meaning. In physics, when a force acts on an object and causes energy to be transferred or transformed, then work is being done on the object. For example, if a weightlifter applies a force to a barbell to lift it, then work has been done on the barbell; chemical energy within the weightlifter's body has been transformed into the gravitational potential energy of the barbell (see Figure 12.1.2).

FIGURE 12.1.2 As a weightlifter lifts a barbell, chemical energy is transformed into gravitational potential energy.

Quantifying work

Work causes a change in energy, i.e. $W = \Delta E$.

More specifically, work is defined as the product of the force causing the energy change and the displacement of the object in the direction of the force during the energy change:

$W = Fs$

where W is work (in J)

F is force (in N)

s is the displacement in the direction of the force (in m).

Since work corresponds to a change in energy, the SI unit of work is also the joule. The definition of work allows us to find a value for a joule in terms of other SI units.

Since $W = Fs$, $1\ \text{J} = 1\ \text{N} \times 1\ \text{m} = 1\ \text{N m}$.

A joule is equal to a newton-metre, that is, a force of 1 N acting over a distance of 1 m does 1 J of work.

Using the definition of a newton:

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

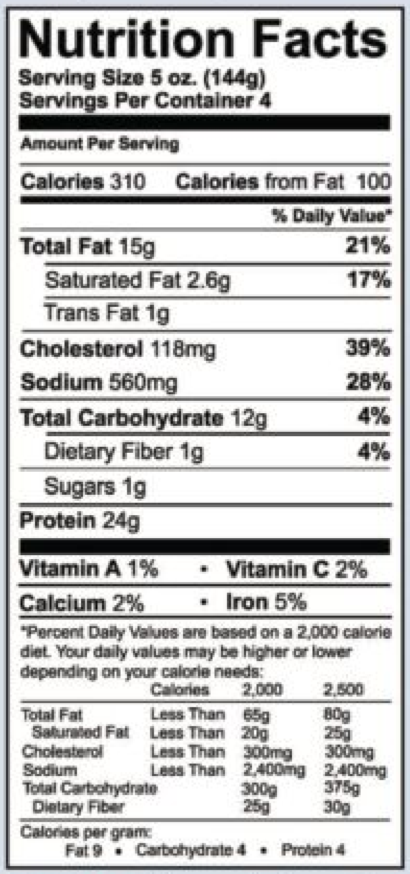
This defines a joule in terms of fundamental units.

Although both force and displacement are vectors, work is a scalar unit. So like energy, work has no direction.

PHYSICSFILE

Units of energy

A number of non-SI units for energy are still in use. When talking about the energy content of food, it is common to use a unit called a calorie (cal) (see Figure 12.1.3). One calorie is defined as the amount of heat required to increase the temperature of 1 g of water by 1°C. This equates to 4.2 J.



Nutrition Facts

Serving Size 5 oz. (144g)
Servings Per Container 4

Amount Per Serving

Calories 310 **Calories** from Fat 100

	% Daily Value*
Total Fat 15g	21%
Saturated Fat 2.6g	17%
Trans Fat 1g	
Cholesterol 118mg	39%
Sodium 560mg	28%
Total Carbohydrate 12g	4%
Dietary Fiber 1g	4%
Sugars 1g	
Protein 24g	

Vitamin A 1% • **Vitamin C** 2%

Calcium 2% • **Iron** 5%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories	2,000	2,500
Total Fat	Less Than	65g	80g
Saturated Fat	Less Than	20g	25g
Cholesterol	Less Than	300mg	300mg
Sodium	Less Than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Calories per gram:
Fat 9 • Carbohydrate 4 • Protein 4

FIGURE 12.1.3 The amount of energy in a serving of food is often measured in calories.

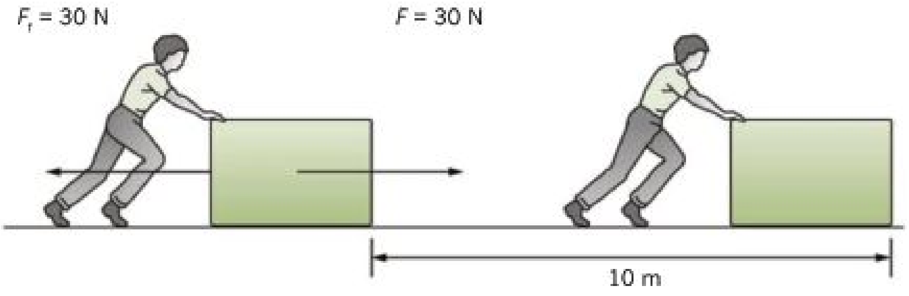
Electrical energy in the home is often measured in kilowatt-hours (kW h). A kilowatt-hour is a very large unit of energy: 1 kW h = 3 600 000 J or 3.6 MJ.

Another, not-so-common unit of energy is the erg (from the Greek word *ergon* for energy). An erg is a very small unit of energy: 1 erg = 10^{-7} J.

Worked example 12.1.1

CALCULATING WORK

A person pushes a heavy box along the ground for 10 m with a force of 30 N. Calculate the amount of work done.



Thinking	Working
Recall the definition of work.	$W = Fs$
Substitute in the values for this situation.	$W = 30 \times 10$
Solve the problem, giving an answer with appropriate units.	$W = 300$ J

Worked example: Try yourself 12.1.1

CALCULATING WORK

A person pushes a heavy wardrobe from one room to another by applying a force of 50 N for a distance of 5 m. Calculate the amount of work done.

Work and friction

The energy change produced by work is not always obvious. Consider Worked example 12.1.1, where 300 J of work was done on a box when it was pushed 10 m. A number of energy outcomes are possible for this scenario.

- In an ideal situation, where there was no friction, all of this work would be transformed into kinetic energy and the box would end up with a higher velocity than before it was pushed.
- In most real situations, where there is friction between the box and the ground, some of the work done would become heat and sound due to friction and the rest would become kinetic energy.
- In the limiting situation, where the force applied is exactly equal to the friction, the box would slide at a constant speed. This means that its kinetic energy would not change, so all of the work done would be converted into heat and sound due to friction.

The displacement of a body is dependent on overcoming the force of friction.

A force with no work

The mathematical definition of work has some unusual implications. One is that if a force is applied to an object but the object does not move, then no work is done.

This appears counterintuitive, that is, it goes against what you would probably expect. An example of this is shown in Figure 12.1.4. While picking up a heavy box requires work, holding the box at a constant height does no work on the box.

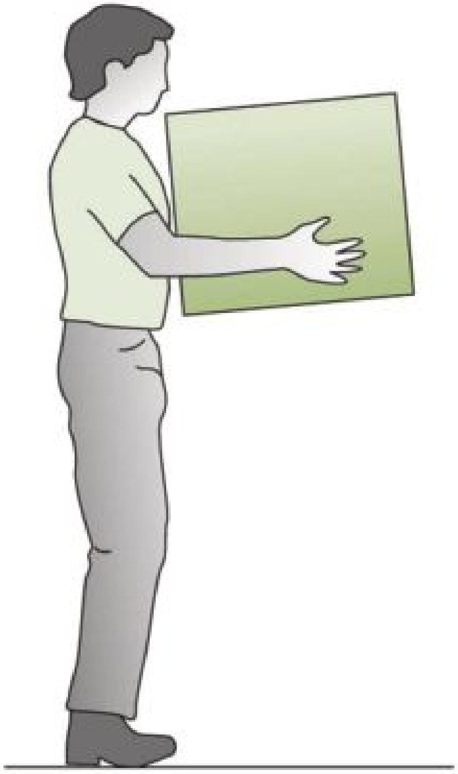


FIGURE 12.1.4 According to the definition of work, no work is done when a person holds a box at a constant height.

Assuming the box has a weight of 100 N and that it is lifted from the ground to a height of 1.2 m, the work done lifting it would be: $W = Fs = 100 \times 1.2 = 120$ J. In this case, energy is being transformed from chemical energy inside the person's body into the gravitational potential energy of the box.

However, when the box is held at a constant height, the definition of work gives: $W = Fs = 100 \times 0 = 0$ J. So, no work is being done *on* the box. Although there would be energy transformations going on inside the person's body to keep their muscles working, the energy of the box does not change, therefore no work has been done on the box.

Work is done only if the net force causes a movement of one body in relation to other bodies.

Work and displacement at an angle

Sometimes, when a force is applied, the object does not move in the same direction as the force. For example, in Figure 12.1.5, when a person pushes a pram, the direction of the force is at an angle downwards, although the pram moves horizontally forwards.

FIGURE 12.1.5 When a person pushes a pram, the force is applied at an angle to the displacement of the pram.

In this case, only the *horizontal component* of the push contributes to the work being done on the pram. The vertical component of this force pushes the pram downwards and is balanced by the normal reaction force from the ground.

In situations like this, work can be calculated using the general equation:

$W = Fs \cos \theta$

where θ is the angle between the force, F, and the displacement, s.

EXTENSION

Resolving forces

In the situation of a person pushing a pram, the general equation $W = Fs$ applies. The person's push can be resolved into a vertical component, $F \sin \theta$, and a horizontal component, $F \cos \theta$, (see Figure 12.1.6). Substituting the horizontal component into the general definition for work gives:

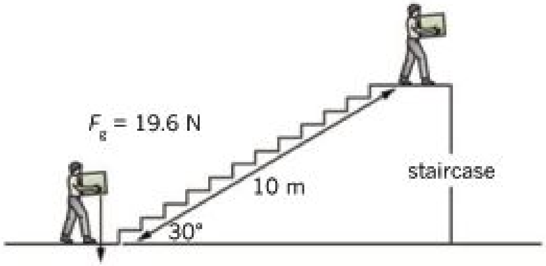
$$
\begin{aligned}
W &= F \cos \theta \times s \\
&= Fs \cos \theta
\end{aligned}
$$

FIGURE 12.1.6 The force applied by the person pushing the pram can be resolved into a horizontal force and a vertical force.

Worked example 12.1.2

WORK WITH FORCE AND DISPLACEMENT AT AN ANGLE

A person carries a box weighing 19.6 N up a 10 m flight of stairs. Calculate the work done against gravity on the box.

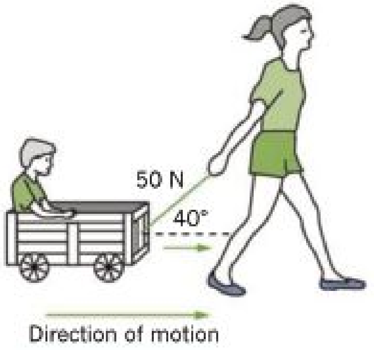


Thinking	Working
Determine values for F, s and θ. Note that the required component of the force is upwards, so the angle is not 30°. It is 90 – 30 = 60°.	Force applied to the box by the person: $F = 19.6$ N upwards Displacement: $s = 10$ m Angle between the force and displacement: $\theta = 60°$
Recall the work equation.	$W = Fs \cos \theta$
Substitute values into the work equation.	$W = 19.6 \times 10 \times \cos 60°$
State the answer with the correct units.	$W = 98$ J

Worked example: Try yourself 12.1.2

WORK WITH FORCE AND DISPLACEMENT AT AN ANGLE

A girl pulls her brother along in a trolley for a distance of 30 m, as shown. Calculate the work done on the box. Give your answer correct to three significant figures.



FORCE–DISPLACEMENT GRAPHS

As its name suggests, a force–displacement graph illustrates the way a force changes with displacement. For a situation where the force is constant, this graph is simple. For example, in Figure 12.1.7, the force–displacement graph for a person picking up a box is a flat horizontal line showing that the force applied to the box is constant throughout the lift.

In contrast, an **elastic** object such as a spring obeys a relationship known as Hooke's law. Hooke's law is discussed in more detail in Section 12.2. Briefly, the law describes how, the more you stretch a spring, the greater the force required to keep stretching it. The force–displacement graph for a spring is also a straight line, but this line shows the direct relationship described by Hooke's law (see Figure 12.1.8). (Note: sometimes, you will see force–displacement graphs for elastic objects labelled as force–extension graphs. In this context, the term extension is the same as displacement.)

Many everyday materials are only partially elastic. Their force–displacement graphs are relatively complex. For example, the force–displacement graph in Figure 12.1.9 for a sports shoe shows that the shoe is close to elastic for low displacements, but at high displacements the restoring force is relatively constant.

FIGURE 12.1.9 The force–displacement graph for a sports shoe is not a straight line; the change in force varies with how much the shoe has been stretched.

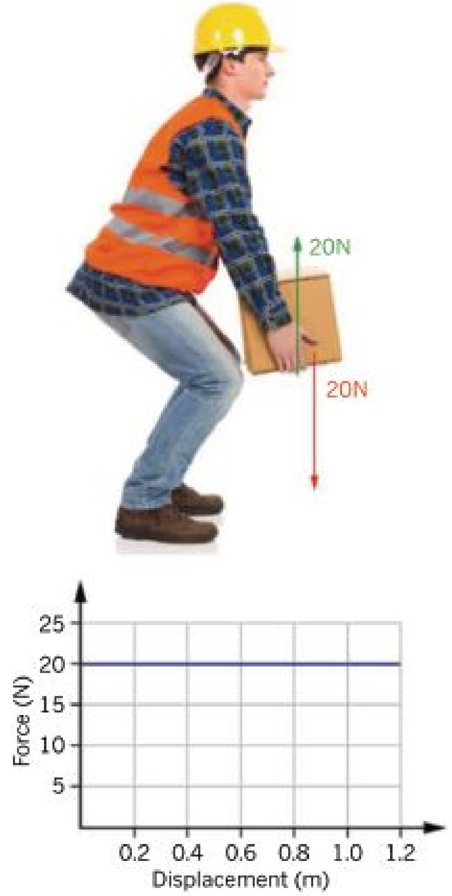


FIGURE 12.1.7 The force–displacement graph for a person picking up a box is a straight, horizontal line, indicating that the force applied to the box by the man is constant throughout the process.

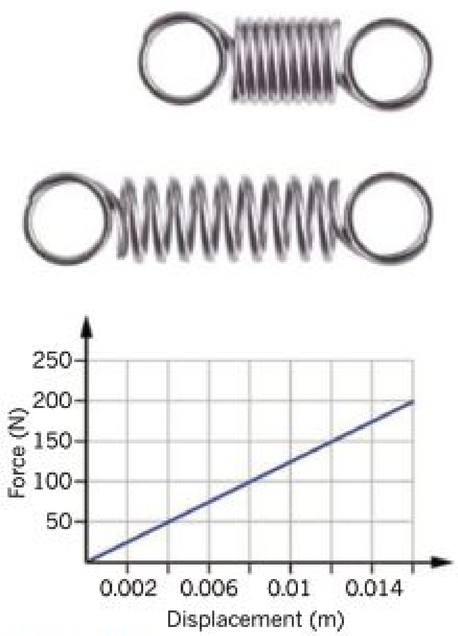


FIGURE 12.1.8 As a spring stretches, more force is required to keep stretching it. The force is proportional to the extension.

CALCULATING WORK FROM A FORCE-DISPLACEMENT GRAPH

When a force changes with displacement, the amount of work done by the force can be calculated from the area under its force–displacement graph.

For a constant force, this is very simple. Considering the earlier example of a person lifting a box, the area can be found by counting the number of 'force times distance' squares under the line. In the example in Figure 12.1.10 there are 6 × 4 = 24 of these squares. Since each square has an area of 5 N × 0.2 m = 1 J, the total work done is 24 J. Alternatively, this area could be found by recognising that it is a rectangle and multiplying length by width to find the area. For Figure 12.1.10, this is 20 N × 1.2 m = 24 J. Note that this second method is exactly the same as using the formula for work: $W = Fs = 20 \times 1.2 = 24$ J. This relationship works in this case because the force is constant.

Similar strategies, either counting grid squares or calculating the area of the shape under the graph, can also be used when the force varies with the displacement.

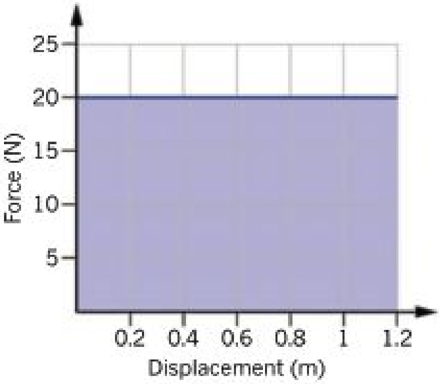
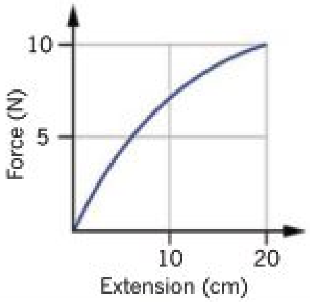


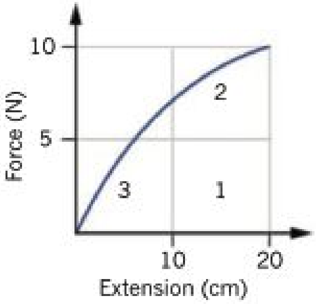
FIGURE 12.1.10 The area under a force–displacement graph gives the work done by the force.

Worked example 12.1.3

WORK FROM THE AREA UNDER A FORCE–DISPLACEMENT GRAPH

Use the force–extension graph for an elastic band to estimate how much work is done in stretching the elastic band 20 cm. Give your answer to the nearest 0.5 J.

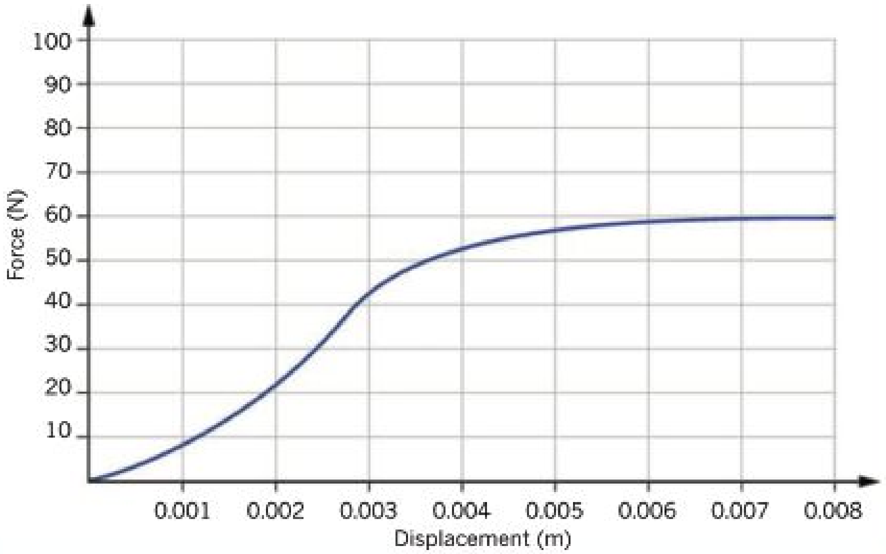


Thinking	Working
Calculate the work value of each grid square.	The dimensions of a grid square are: Force: 5 N, Extension: 10 cm = 0.1 m Area of 1 square = 5 × 0.1 = 0.5 J
Count the number of grid squares under the curve.	Only count grid squares that are more than half under the curve. If the curve cuts a square in half, count every second one.  Number of squares = 3
Multiply the number of grid squares under the curve by the work value of each grid square.	$W = 3 \times 0.5$ J = 1.5 J

Worked example: Try yourself 12.1.3

WORK FROM THE AREA UNDER A FORCE–DISPLACEMENT GRAPH

While jogging, a person's shoes stretch by an average of 3 mm with each step. Use the force–displacement graph for a sports shoe to estimate how much work is done on the shoe with each step. Give your answer to the nearest 0.01 J.



12.1 Review

SUMMARY

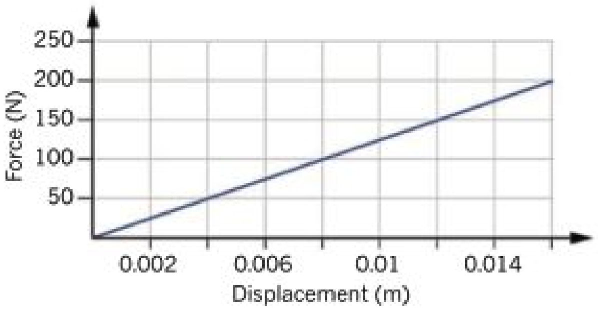
- Energy is the capacity to cause a change.
- Energy is conserved. It can be transferred or transformed, but not created or destroyed.
- There are many different forms of energy. These can be broadly classified as either kinetic (associated with movement) or potential (associated with the relative positions of objects).
- Work is done when energy is transferred or transformed.
- Work is done when a force causes an object to be displaced.
- Work is the product of force and displacement: $W = Fs$
- When a force produces no displacement, or when the force and displacement are at right angles to each other, no work is done.
- Work is equal to the area under a force–displacement graph.
- A straight horizontal line in a force–displacement graph represents a constant force.
- The relationship between force and displacement for an elastic object is represented as a straight diagonal line in a force–displacement graph.

KEY QUESTIONS

1 When accelerating at the beginning of a ride, a cyclist applies a force of 500 N for a distance of 20 m. What is the work done by the cyclist on the bike?

2 In the case of a person leaning on a solid brick wall, explain why no work is being done.

3 A spring with this force–displacement graph is stretched as shown. Using the formula for the area of a triangle, calculate the work done to stretch the spring by 0.015 m.



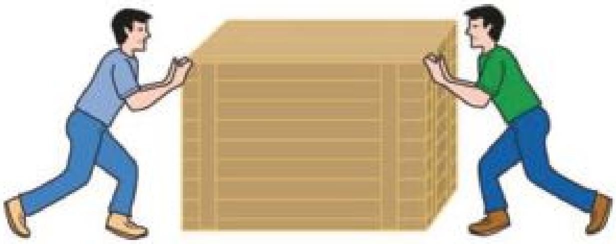
12.1 Review *continued*

4 A cyclist does 2700 J of work when she rides her bike at a constant speed for 150 m. Calculate the average force the cyclist applies over this distance.

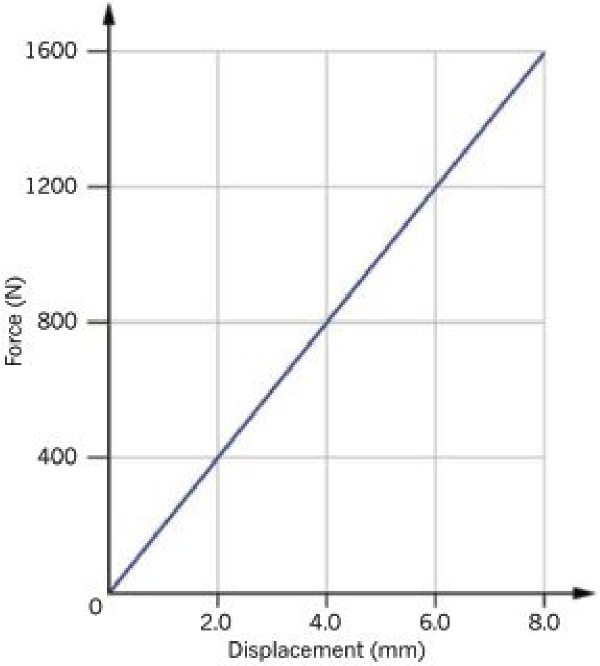
5 A rope at 40° to the horizontal is used to drag a heavy box along the ground for a distance of 5.0 m. Calculate the work done if the tension in the rope is 80 N. Give your answer correct to the nearest 10 J.

6 Explain why the equation $W = Fs$ cannot be used to calculate the work done in compressing a spring.

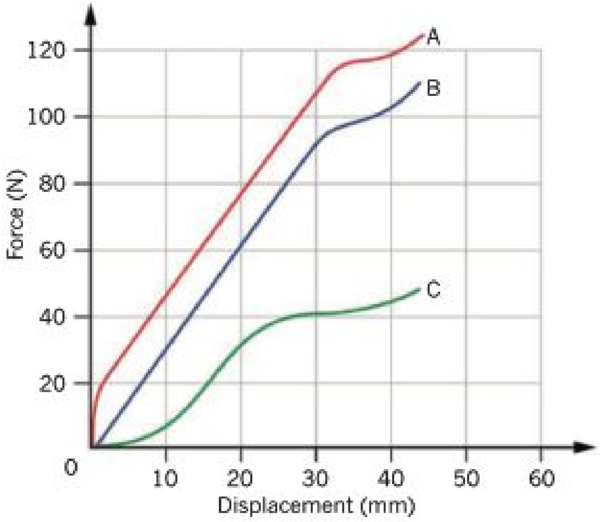
7 Two people push in opposite directions on a heavy box. One person applies 50 N of force, the other applies 40 N of force. There is 10 N of friction between the box and the floor which means that the box does not move. What is the work done by the person applying 50 N of force?



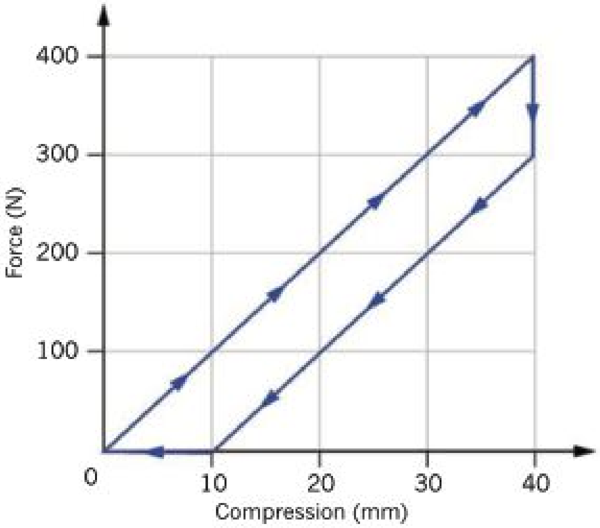
8 The strings of a graphite-head tennis racquet have the force–displacement graph shown. Calculate the work done when the strings displace by 6 mm.



9 Three different springs have the force–displacement graphs shown. Estimate the work done by stretching each of the springs by 40 mm. Give your answers correct to two significant figures.



10 The following diagram is a simplified representation of the forces acting on a basketball when it bounces. The upwards arrows show the force as the basketball compresses and the downwards arrows show the force as it rebounds.



a Calculate the work done on the basketball when it compresses by 40 mm.

b Calculate the work done by the ball as it decompresses from a compression of 40 mm.

c Explain why your answers to a and b are different.

12.2 Mechanical energy

Mechanical energy is the energy that a body possesses due to its position or motion. Kinetic energy, gravitational energy and elastic potential energy are all forms of mechanical energy.

Any object that moves, such as those shown in Figure 12.2.1, has kinetic energy. Many real-life energy interactions, like throwing a ball, involve objects with kinetic energy. Some of these, like car collisions, have life-threatening implications. Hence, it is important to be able to quantify (i.e. find numerical values for) the kinetic energy of an object.

One of the easiest forms of potential energy to study is gravitational potential energy. Any object that is lifted above the Earth's surface has the capacity to cause change due to its position in the Earth's gravitational field. An understanding of gravitational potential energy is essential to understanding common energy transformations.

FIGURE 12.2.1 Any moving object, regardless of its size, has kinetic energy.

THE KINETIC ENERGY EQUATION

Kinetic energy is the energy of motion. It can be quantified by calculating the amount of work needed to give an object its velocity.

Consider the dynamics cart in Figure 12.2.2 of mass, m, starting at rest (i.e. $u = 0$). It is pushed with force, F, which acts while the cart undergoes a displacement, s, and gains a final velocity, v.

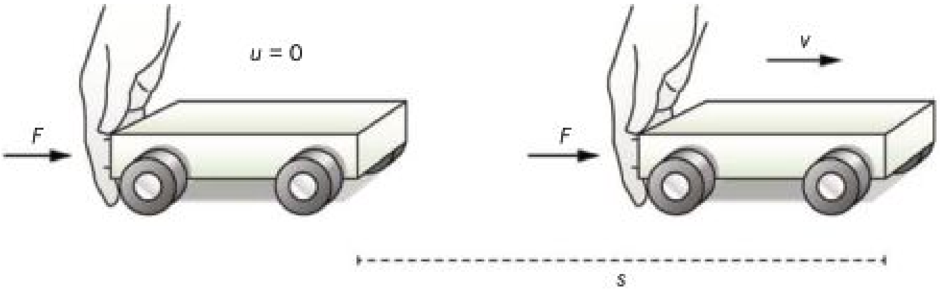


FIGURE 12.2.2 The kinetic energy of a dynamics cart can be calculated by considering the force (F) acting on it over a given displacement (s).

The work done by the force, W, causes a change in kinetic energy from its initial value $\frac{1}{2}mu^2$ to a new value of $\frac{1}{2}mv^2$.

The relationship between the work done and the change in kinetic energy can be written mathematically as:

$$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

where W is work (in J)

m is mass (in kg)

u is initial velocity (in m s^{-1})

v is final velocity (in m s^{-1}).

This equation is known as the 'work–energy theorem'.

In this situation, the cart was originally at rest (i.e. $u = 0$) so:

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

Assuming that no energy was lost as heat or noise and that all of the work is converted into kinetic energy, this equation gives us a mathematical definition for the kinetic energy of the cart in terms of its mass and velocity:

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

where E_k is kinetic energy (in J)

EXTENSION

Expressing the amount of work

Considering the scenario described in Figure 12.2.2, the work done by the force is given by the equation $W = Fs$. The force causes the cart to accelerate according to Newton's second law, $F = ma$.

Rearranging the equation of motion $v^2 = u^2 - 2as$ gives:

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

Combining this with $F = ma$ means that the force acting on the cart can be given by the equation:

$$F = m\left(\frac{v^2 - u^2}{2s}\right)$$

This equation can be transposed to find an expression for the amount of work (Fs) done on the cart:

$$F = \frac{m}{2}(v^2 - u^2)$$

$$Fs = \frac{m}{2}(v^2 - u^2)$$

$$= F = \frac{1}{2}m(v^2 - u^2)$$

Since $W = Fs$:

$$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

Worked example 12.2.1

CALCULATING KINETIC ENERGY

A car with a mass of 1200 kg is travelling at 90 km h^{-1}. Calculate its kinetic energy at this speed.	
Thinking	**Working**
Convert the car's speed to m s^{-1}.	$90 \text{ km h}^{-1} = \frac{90 \text{ km}}{1 \text{ h}} = \frac{90000 \text{ m}}{3600 \text{ s}}$ $= 25 \text{ m s}^{-1}$
Recall the equation for kinetic energy.	$E_k = \frac{1}{2}mv^2$
Substitute the values for this situation into the equation.	$E_k = \frac{1}{2} \times 1200 \times 25^2$
State the answer with appropriate units.	$E_k = 375000 \text{ J} = 375 \text{ kJ}$

Worked example: Try yourself 12.2.1

CALCULATING KINETIC ENERGY

A person crossing the street is walking at 5.0 km h^{-1}. If the person has a mass of 80 kg, calculate their kinetic energy. Give all answers correct to two significant figures.

APPLYING THE WORK–ENERGY THEOREM

The work–energy theorem can be seen as a definition for the *change* in kinetic energy produced by a force:

$$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2 = (E_k)_{\text{final}} - (E_k)_{\text{initial}} = \Delta E_k$$

Worked example 12.2.2

CALCULATING KINETIC ENERGY CHANGES

A 2 tonne truck travelling at 100 km h^{-1} slows to 80 km h^{-1} before turning a corner.

a Calculate the work done by the brakes to make this change. Give answers to two significant figures.

Thinking	Working
Convert the values into SI units.	$u = 100 \text{ km h}^{-1} = \frac{100 \text{ km}}{1 \text{ h}} = \frac{100000 \text{ m}}{3600 \text{ s}}$ $= 28 \text{ m s}^{-1}$ $v = 80 \text{ km h}^{-1} = \frac{80 \text{ km}}{1 \text{ h}} = \frac{90000 \text{ m}}{3600 \text{ s}}$ $= 22 \text{ m s}^{-1}$ $m = 2$ tonne $= 2000$ kg
Recall the work–energy theorem.	$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$
Substitute the values for this situation into the equation.	$W = \frac{1}{2}(2000 \times 22^2) - \frac{1}{2}(2000 \times 28^2)$
State the answer with appropriate units.	$W = -300000 \text{ J} = -300 \text{ kJ}$ Note: the negative value indicates that the work has caused kinetic energy to decrease.

b If it takes 50 m for this deceleration to take place, calculate the average force applied by the truck's brakes.

Thinking	Working
Recall the definition of work.	$W = Fs$
Substitute the values for this situation into the equation. Note: The negative has been ignored since work is a scalar.	$300000 \text{ J} = F \times 50 \text{ m}$
Transpose the equation to find the answer.	$F = \frac{W}{s} = \frac{30000 \, J}{50 \, m} = 6000 \text{ N}$

Worked example: Try yourself 12.2.2

CALCULATING KINETIC ENERGY CHANGES

As a bus with a mass of 10 tonnes approaches a school it slows from 60 km h^{-1} to 40 km h^{-1}.

a Calculate the work done by the brakes in the bus. Give answers to two significant figures.

b The bus travels 40 m as it decelerates. Calculate the average force applied by the truck's brakes.

Notice that the definitions for kinetic energy and change in kinetic energy have been derived entirely from known concepts: the definition of work, Newton's second law and the equations of motion. This makes kinetic energy appear a redundant concept. However, using kinetic energy calculations can often make analysis of physical interactions quicker and easier, particularly in situations where acceleration is not constant.

Worked example 12.2.3

CALCULATING SPEED FROM KINETIC ENERGY

The engine of a 1400 kg car can do 900 kJ of work in 10 s. Assuming all of this work is converted into kinetic energy, calculate the speed of the car after this time in km h^{-1}. Give your answer correct to two significant figures.

Thinking	Working
Recall the equation for kinetic energy.	$E_k = \frac{1}{2}mv^2$
Transpose the equation to make v the subject.	$v = \sqrt{\frac{2E_k}{m}}$
Substitute the values for this situation into the equation.	$v = \sqrt{\frac{2 \times 900 \times 10^3}{1400}} = 36 \text{ m s}^{-1}$
State the answer with appropriate units.	$v = 36 \times 3.6 = 130 \text{ km h}^{-1}$

Worked example: Try yourself 12.2.3

CALCULATING SPEED FROM KINETIC ENERGY

A 300 kg motorbike has 150 kJ of kinetic energy. Calculate the speed of the motorbike in km h^{-1}. Give your answer correct to two significant figures.

DEFINING GRAVITATIONAL POTENTIAL ENERGY

Gravitational potential is a measure of the amount of energy available to an object due to its position in a gravitational field. The gravitational potential energy of an object can be calculated from the amount of work that must be done against gravity to get the object into its position.

Consider the weightlifter lifting a barbell in Figure 12.2.3. Assuming that the bar is lifted at a constant speed, then the weightlifter must apply a lifting force equal to the force due to gravity on the barbell, F_g. The lifting force, F_l, is applied over a displacement, Δh, corresponding to the change in height of the barbell.

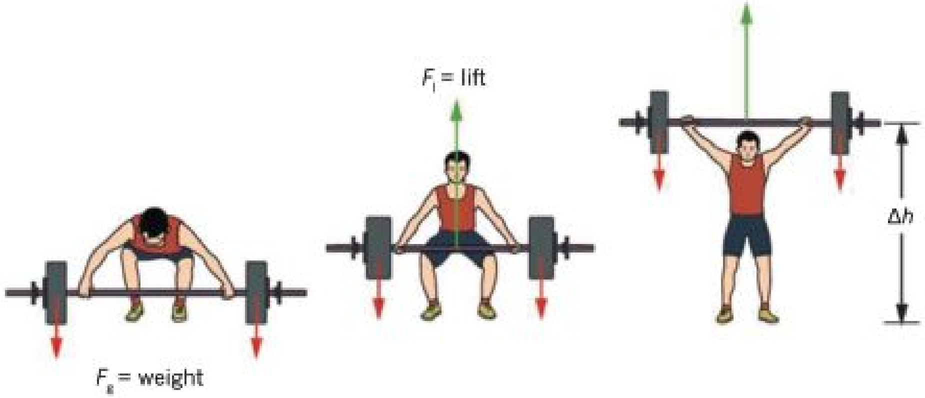


FIGURE 12.2.3 A weightlifter applies a constant force over a fixed distance to give the barbell gravitational potential energy.

The work done against gravity by the weightlifter is:

$W = Fs = F_g \Delta h$

Since the force due to gravity $F_g = mg$, the work done can be written as:

$W = mg\Delta h$

The work carried out in this example has resulted in the transformation of chemical energy within the weightlifter into gravitational potential energy. Since the change in gravitational potential energy of the barbell is:

$\Delta E_g = mg\Delta h$

Taking the ground as the point where the gravitational potential energy is zero (that is, $E_g = 0$) the gravitational potential energy of an object, due to the work done against a gravitational field, is given by:

$E_g = mg\Delta h$

where E_g is the gravitational potential energy (in J)

m is the mass of the object (in kg)

g is the gravitational field strength (9.8 N kg^{-1} on Earth)

Δh is the change in height of the object (in m).

Worked example 12.2.4

CALCULATING GRAVITATIONAL POTENTIAL ENERGY

A weightlifter lifts a barbell which has a total mass of 80 kg from the floor to a height of 1.8 m above the ground. Calculate the gravitational potential energy of the barbell at this height. Give your answer correct to two significant figures.

Thinking	Working
Recall the formula for gravitational potential energy.	$E_g = mg\Delta h$
Substitute the values for this situation into the equation.	$E_g = 80 \times 9.8 \times 1.8$
State the answer with appropriate units and significant figures.	$E_g = 1411.2$ J = 1.4 kJ

Worked example: Try yourself 12.2.4

CALCULATING GRAVITATIONAL POTENTIAL ENERGY

A person doing their grocery shopping lifts a 5 kg grocery bag to a height of 30 cm. Calculate the gravitational potential energy of the grocery bag at this height. Give your answer correct to two significant figures.

GRAVITATIONAL POTENTIAL ENERGY AND REFERENCE LEVEL

When calculating gravitational potential energy, it is important to carefully define the level that corresponds to $E_g = 0$. Often this can be taken to be the ground or sea level, but the zero potential energy reference level is not always obvious.

It does not really matter which point is taken as the zero potential energy reference level, as long as the chosen point is used consistently throughout a particular problem (see Figure 12.2.4). If objects move below the reference level, then their energies will become negative and should be interpreted accordingly.

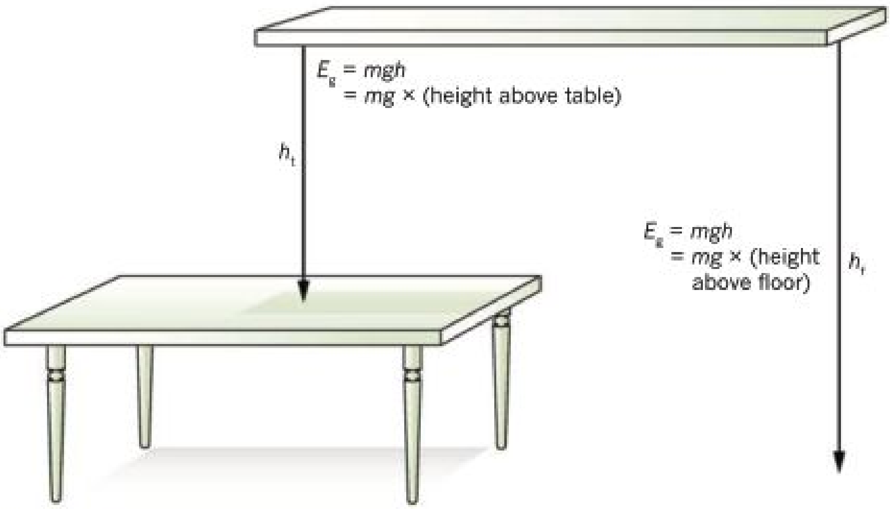


FIGURE 12.2.4 In this situation, the zero potential energy reference point could be taken as either the level of the table or the floor.

Worked example 12.2.5

CALCULATING GRAVITATIONAL POTENTIAL ENERGY RELATIVE TO A REFERENCE LEVEL

A weightlifter (m = 60 kg) lifts a 50 kg bar through a distance of 40 cm. Calculate the increase in gravitational potential energy of the bar with each lift. Use g = 9.8 N kg^{-1} and state your answer correct to three significant figures.

Thinking	Working
Recall the formula for gravitational potential energy.	$E_g = mg\Delta h$
Identify the relevant values for this situation. Only the mass of the bar is being lifted (the weightlifter's mass is a distractor). Take the weightlifter's body as the zero potential energy level.	m = 50 kg g = 9.8 N kg^{-1} Δh = 40 cm = 0.4 m
Substitute the values for this situation into the equation.	$E_g = 50 \times 9.8 \times 0.4$
State the answer with appropriate units and significant figures.	E_g = 196 J

Worked example: Try yourself 12.2.5

CALCULATING GRAVITATIONAL POTENTIAL ENERGY RELATIVE TO A REFERENCE LEVEL

A father picks up his baby from its bed. The baby has a mass of 6.0 kg and the mattress of the bed is 70 cm above the ground. When the father holds the baby in his arms, it is 125 cm off the ground. Calculate the increase in gravitational potential energy of the baby, taking g as 9.8 N kg^{-1} and giving your answer correct to two significant figures.

EXTENSION

The high jump

Science has long been used in sport to help athletes gain a competitive edge. The concept of gravitational potential energy is of obvious importance to a high jumper. Clearly, the high jumper must do enough work in their jump to create sufficient gravitational potential energy to clear the bar.

The modern high jump technique known as the Fosbury flop gets the high jumper to bend their body as they go over the bar. This is illustrated in Figure 12.2.5.

FIGURE 12.2.5 In the Fosbury flop technique, a high jumper must bend their body over the bar.

When the technique is correctly performed, most of the mass of the jumper (e.g. their head, arms and legs) is actually lower than the bar throughout the jump. In other words, the centre of mass of the jumper passes below the bar while their body bends over it, as shown in Figure 12.2.6.

FIGURE 12.2.6 The path of the centre of mass of the high jumper (shown by the dashed curve) passes below the high jump bar.

If the jumper's technique is right, the high jumper does not have to produce enough gravitational potential energy to lift their body above the bar. Without this technique, the world records for this event would probably be much lower than their current marks.

PHYSICSFILE

Newton's universal law of gravitation

The formula $E_g = mg\Delta h$ is based on the assumption that the Earth's gravitational field is constant. Newton's universal law of gravitation predicts that the Earth's gravitational field will decrease with altitude. However, this decrease only becomes significant many kilometres above the Earth's surface. For everyday purposes, the assumption of a constant gravitational field is valid.

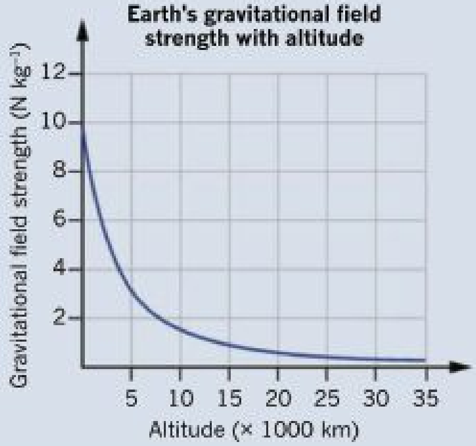


FIGURE 12.2.7 The Earth's gravitational field strength decreases with altitude.

ELASTIC MATERIALS AND ELASTIC POTENTIAL ENERGY

The third aspect of mechanical energy under consideration here is **elastic potential energy**. Like gravitational potential energy, it occurs in situations where energy can be considered to be stored temporarily. When this energy is released, work may be done on an object.

Elastic potential energy is stored when a spring is stretched, a rubber ball is squeezed, air is compressed in a tyre or a bungee-jumper's rope is extended during a jump. Since each object possesses energy due to its position or motion, all of these situations suit the earlier definition of mechanical energy.

Materials that have the ability to store elastic potential energy when work is done on them, and then release this energy, are called *elastic* materials. Metal springs and bouncing balls are common examples; however, many other materials are at least partially elastic. If their shape is manipulated, items such as our skin, metal hair clips and wooden rulers all have the ability to restore themselves to their original shape once released.

Materials that do not return to their original shape and release their stored potential energy as mechanical energy are referred to as plastic materials. Plasticine is an example of a very plastic material.

Ideal springs obey Hooke's law

Springs are very useful items in everyday life due to the consistent way in which many of them respond to forces and store energy. When a spring is stretched or compressed by an applied force, elastic potential energy is being stored. In order to store this energy, work must be done on the spring.

Recall that if a constant force is applied to an object, and a displacement occurs in the direction of that force, then the quantity of work done can be calculated using $W = Fs$.

This formula can therefore be used when a constant force, F, has been applied to a spring and a given compression or extension, Δx, occurs. However, it is more interesting to examine how a spring will behave under a range of conditions.

Consider the situation in which a spring is stretched by the application of a steadily increasing force. As the force increases, the extension of the spring, Δx, can be graphed against the applied force, F. Well-designed springs will extend in proportion to the applied force, when the load is not too large. For example, if a 10 N force produced an extension of 6 cm, then a 20 N force would produce an extension of 12 cm. For ideal springs the resulting graph of applied force versus extension would be linear, as shown in Figure 12.2.8.

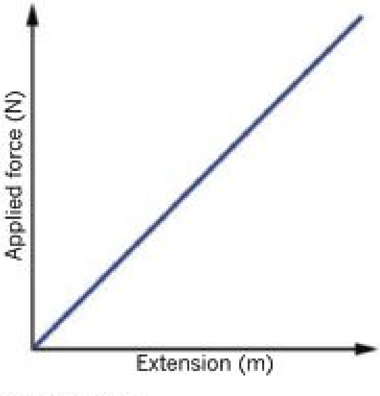


FIGURE 12.2.8 Ideal springs obey Hooke's law.

The gradient of a force–extension graph tells us the force, in newtons, required to produce each unit of extension in metres.

The gradient of the graph is called the **spring constant**, k, and is measured in N m^{-1}. The gradient indicates the stiffness of the spring. For an ideal spring this gradient has a set value (as the F–Δx graph has a constant slope). A very stiff spring that is difficult to stretch would have a steep gradient; that is, a large value of k.

Although k is usually called the spring constant, it is sometimes called the stiffness constant or force constant of a spring. A spring constant of $k = 1500\ \text{N m}^{-1}$ indicates that for every metre that the spring is stretched or compressed, a force of 1500 N is required. This does not necessarily mean that the spring can be stretched by 1 m, but it indicates that the force and the change in length are in this proportion.

The relationship between the applied force and the subsequent extension or compression of an ideal spring is known as **Hooke's law**. For ideal springs, $F \propto \Delta x$ or $F = k\Delta x$.

However, when using the energy stored by stretched or compressed springs it is appropriate to refer to the force that the distorted spring is able to exert (rather than the force that was applied to it). Newton's third law says that an extended or compressed spring in equilibrium is able to exert a restorative force equal in size but opposite in direction to the force that is being applied to it. Therefore Hooke's law is often written in the form shown below.

Hooke's law states that the force applied by a spring is directly proportional, but opposite in direction, to the spring's extension or compression. That is:

$F = -k\Delta x$

where F is the force applied by the ideal spring (in N)

k is the spring constant (force constant or stiffness constant) (in N m^{-1})

Δx is the amount of extension or compression of the ideal spring (in m).

Note: the negative sign in Hooke's law indicates that the restorative force inside the spring and the extension are in opposite directions.

CALCULATING ELASTIC POTENTIAL ENERGY

Work must be done in order to store elastic potential energy, E_s, in any elastic material. Essentially, the energy is stored within the atomic bonds of the material as it is compressed or stretched. The amount of elastic potential energy stored is given by the area under the force–extension graph for the item.

For materials that obey Hooke's law (as seen in Figure 12.2.8), an expression can be derived for the area under the force–extension graph.

Work done = area under the F–Δx graph

W = area of triangle

$= \frac{1}{2} b \times h$

As Δx is the base, b, and F is the height, h:

$W = \frac{1}{2}\Delta x \times F$

But $F = k\Delta x$ so:

$W = \frac{1}{2}\Delta x \times k\Delta x$

$= \frac{1}{2}k\Delta x^2$

$E_s = \frac{1}{2}k\Delta x^2$

The elastic (or spring) potential energy, E_s, stored in an object is given by the area under the force–extension graph for the object. For objects that obey Hooke's law, the spring potential energy is given by:

$E_s = \frac{1}{2}k\Delta x^2$

where E_s is the elastic potential energy stored during the extension/compression (in J)

k is the spring constant (force constant or stiffness constant) (in N m^{-1})

Δx is the amount of extension or compression of the ideal spring (in m).

PHYSICSFILE

Climbing ropes

The ropes used by rock climbers have elastic properties that can save lives during climbing accidents. Ropes that were used in the nineteenth century were made of hemp, which is strong but does not stretch a lot. When climbers using these ropes fell, they stopped very abruptly. The resulting large forces acting on the climbers caused many serious injuries.

Modern ropes are made of a continuous-drawn nylon fibre core and a protective textile covering. They have a slightly lower spring constant than the older style ropes and stretch significantly (up to several metres) when stopping a falling climber. This reduces the stopping force acting on the climber. Ropes with even lower spring constants are suitable for bungee jumping. Rock climbers tend to avoid these ropes—bouncing up and down the rock face is not advisable!

FIGURE 12.2.9 Modern climbing ropes reduce the stopping force on falling climbers, thereby reducing serious injuries.

Worked example 12.2.6

CALCULATING ELASTIC POTENTIAL ENERGY

A spring with a spring constant of 75.0 N m^{-1} is stretched from its original length of 25.0 cm to 32.0 cm. Calculate the elastic potential energy stored in this spring.

Thinking	Working
Identify the variables involved and state them with their directions in their standard form.	$k = 75.0$ N m^{-1} $x_f = 0.320$ m $x_i = 0.250$ m
Determine the extension of the spring.	$\Delta x = x_f - x_i$ $= (0.320 - 0.250)$ $= 0.070$ m
Use the equation for elastic potential energy $E_s = \frac{1}{2}k\Delta x^2$.	$E_s = \frac{1}{2}k\Delta x^2$ $= \frac{1}{2}(75.0)(0.070)^2$ $= 0.184$ J

Worked example: Try yourself 12.2.6

CALCULATING ELASTIC POTENTIAL ENERGY

A spring with a spring constant of 2050 N m^{-1} is stretched from its original length of 45.0 cm to 54.0 cm. Calculate the elastic potential energy stored in this spring.

Although many materials (at least for a small load) extend in proportion to the applied force, many materials have force–extension graphs more like that shown in Figure 12.2.10. For these materials the counting squares method can be used to approximate the elastic potential energy stored.

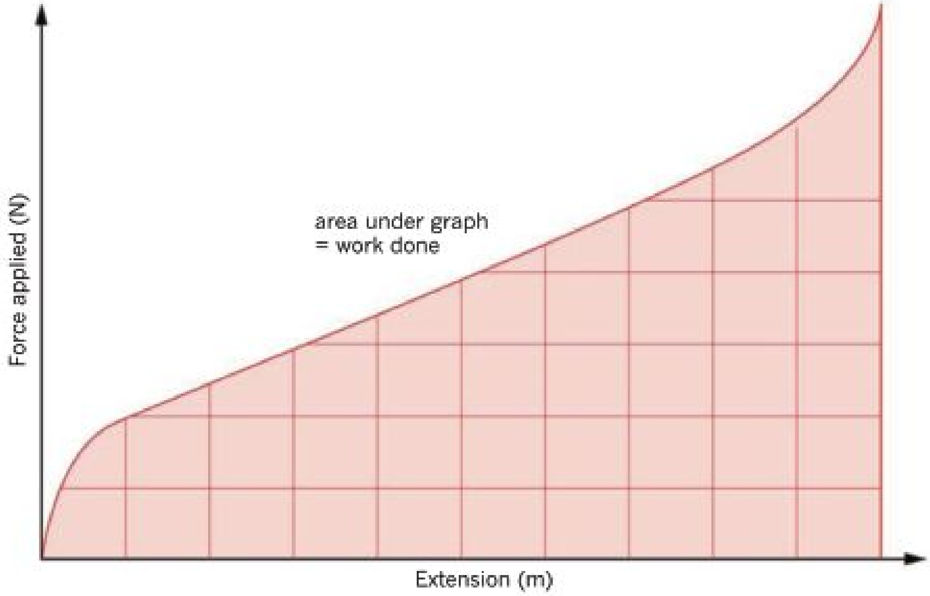


FIGURE 12.2.10 Elastic potential energy is a form of mechanical energy. The area under a graph is the elastic potential energy stored in the object.

PHYSICS IN ACTION

Energy transformations

The world record for the men's pole vault is over 6 m—about as high as a single-storey house! The women's record is just over 5 metres. During the jump, a number of energy transformations take place. The athlete has kinetic energy as she runs in. This kinetic energy is used to bend the pole and carry the athlete forwards over the bar. As the pole bends, energy is stored as elastic potential energy. The athlete uses this stored energy to increase her gravitational potential energy and, hopefully, raise her centre of mass over the bar. Once the pole has been released and the bar has been cleared, the gravitational potential energy of the athlete is transformed into kinetic energy as she falls towards the mat. The energy changes are analysed by making some assumptions about the athlete and the jump.

Assume that the athlete has a mass of 60 kg and runs in at 7.0 m s^{-1}. Treat the athlete as a point mass located at her centre of mass, 1.2 m above the ground. The athlete raises her centre of mass to a height of 5.0 m as she clears the bar, and her speed at this point is just 1.0 m s^{-1}. As she plants the pole in the stop, the pole has not yet been bent and so it has no elastic potential energy. Using:

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

the vaulter's total energy at this point is 2180 J (see Figure 12.2.11).

When the vaulter passes over the bar, the pole is straight again and so has no elastic potential energy. Taking the ground as zero height, and using the same relationship as above, the vaulter's total energy is now 2970 J. This does not seem consistent with the conservation of energy as there is an extra 790 J. The extra energy is from the muscles in her body. Just before the athlete plants the pole, she raises it over her head. Then, after the pole is planted but before she leaves the ground, the athlete uses her arms to bend the pole (see Figure 12.2.12). She pulls downwards on the pole with one arm while the other arm pushes upwards. The effect of these forces is to do work on the pole and store some extra elastic potential energy in it. This work will be converted into gravitational potential energy later in the jump. Energy has also been put into the system by the muscles of the athlete as they do work after she has left the ground. Throughout the jump, she uses her arm muscles to raise her body higher. At the end of the jump, she is actually ahead of the pole and pushing herself up off it. In effect, she has been using the pole to push up off the ground.

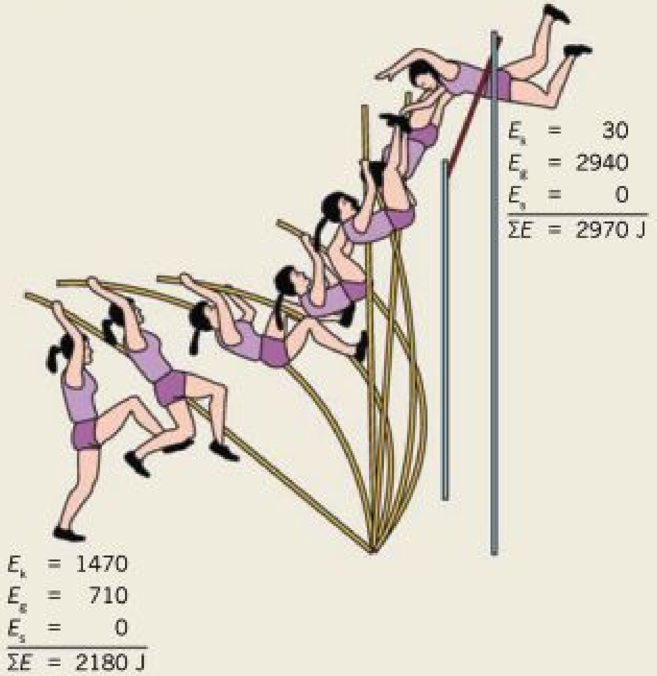


FIGURE 12.2.11 These diagrams, drawn at equal time intervals, indicate that this vaulter slows down as she nears the bar. Her initial kinetic energy is stored as elastic potential energy in the bent pole, and finally transformed into gravitational potential energy and kinetic energy, enabling her to clear the bar.

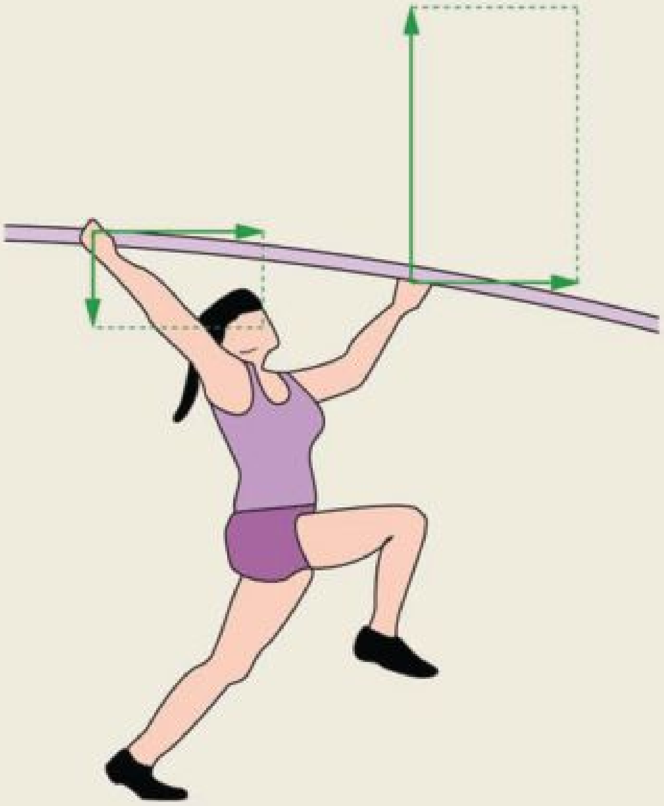


FIGURE 12.2.12 As the pole is planted, the vaulter uses her arms to bend the bar. The forces are shown by the vectors. By bending the bar, the athlete has stored energy which will later be transformed into gravitational potential energy.

12.2 Review

SUMMARY

- All moving objects have kinetic energy.
- The kinetic energy of an object is equal to the work required to accelerate the object from rest to its final velocity.
- The kinetic energy of an object is given by the equation:
  - $E_k = \frac{1}{2}mv^2$
- The work–energy theorem defines work as *change* in kinetic energy:
  - $W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2 = \Delta E_k$
- Gravitational potential energy is the energy an object has due to its position in a gravitational field.
- The gravitational potential energy of an object, E_g, is given by the equation:
  $E_g = mg\Delta h$
- Gravitational potential energy is calculated relative to a zero potential energy reference level, usually the ground or sea level.
- Ideal materials extend or compress in proportion to the applied force; that is they obey Hooke's law:
  $F = -kx$
- The elastic potential energy, E_s, stored in an object is given by the area below a force–extension graph for that object.
- The elastic potential energy, E_s, can be calculated for a spring that obeys Hooke's law using the equation:
  $E_s = \frac{1}{2}k\Delta x^2$

KEY QUESTIONS

1 The mass of a motorbike together with its rider is 230 kg. If the motorbike is travelling at 80 km h^{-1}, calculate its kinetic energy.

2 A 1500 kg car is travelling at 17 m s^{-1}. How much work would its engine need to do to accelerate it to 28 m s^{-1}?

3 A cyclist has a mass of 72 kg and is riding a bicycle which has a mass of 9 kg. When riding at top speed on the bicycle, their kinetic energy is 5 kJ. Calculate the top speed of the cyclist to two significant figures.

4 By how much is kinetic energy increased when the mass of an object is doubled?

5 A 57 g tennis ball is thrown 8.2 m into the air. Use $g = 9.8$ m s^{-2}.

a Calculate the gravitational potential energy of the ball at the top of its flight.

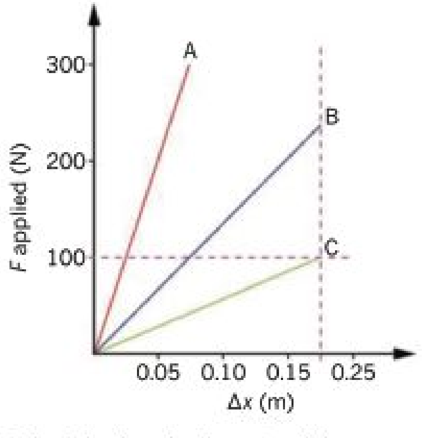
b Calculate the gravitational potential energy of the ball when it has fallen halfway back down to Earth.

6 When climbing Mount Everest ($h = 8848$ m), a mountain climber stops to rest at North Base Camp ($h = 5150$ m). If the mountain climber has a mass of 65.0 kg, how much gravitational potential energy will she gain in the final section of her climb (i.e. from base camp to the summit)? For simplicity, assume that g remains at 9.8 N kg^{-1}.

7 A 30.0 cm ideal spring has a spring constant of 625 N m^{-1}. Calculate the total length of the spring after a weight force of 25.0 N is hung on it.

8 An ideal spring has a spring constant of 320 N m^{-1}. Calculate the elastic potential energy stored in the spring after a weight that is hung on it causes the spring to be 7.00 cm longer.

9 The force–extension graph for three different springs is shown below. The three springs have the following spring constants: $k_A = 4000$ N m^{-1}, $k_B = 1330$ N m^{-1} and $k_C = 500$ N m^{-1}.



Calculate the elastic potential energy stored in each spring when a force of 100 N is applied.

10 A student collects two ideal springs each with a different stiffness. She determines that spring A has the steepest slope on a force–extension graph, while spring B has the flattest slope. What can the student conclude from these results?

12.3 Using energy: Power and efficiency

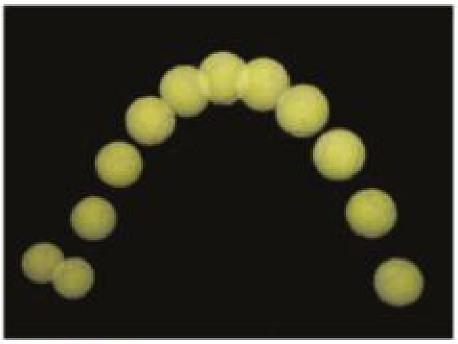


FIGURE 12.3.1 Bouncing tennis ball.

In many situations, energy is transformed between kinetic and gravitational potential energy. For example, when a tennis ball bounces, as shown in Figure 12.3.1, much of its kinetic energy is converted into gravitational potential energy and then back into kinetic energy again.

In analysing this type of situation, the concept of mechanical energy is useful. Mechanical energy is the sum of the potential energies available to an object and the kinetic energy of an object.

In situations where mechanical energy is conserved, it is possible to use this to predict the outcome of the situation. Where mechanical energy is not conserved, this can be used to help identify other important energy transformations.

When considering energy changes, the rate at which work is done is often important. For example, if two cars have the same mass, then the amount of work required to accelerate each car from a standing start to 100 km h^{-1} will be the same. However, the fact that one car can do this more quickly than another may be an important consideration for some drivers when choosing which car to buy.

Physicists describe the rate at which work is done using the concept of power. Like work and energy, this is a word that takes on a specific meaning in a scientific context.

MECHANICAL ENERGY

For falling objects, the mechanical energy is calculated from the sum of its kinetic and gravitational potential energies:

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

This is a useful concept in situations where gravitational potential energy is converted into kinetic energy or vice versa. For example, consider a tennis ball with a mass of 60.0 g that is dropped from a height of 1.00 m. Initially, its total mechanical energy would comprise the kinetic energy, which would be 0 J, and the gravitational potential energy that is stored at this height (taking g = 9.80 m s^{-2}):

$$E_g = mgh = 0.0600 \times 9.80 \times 1.00 = 0.588 \text{ J}$$

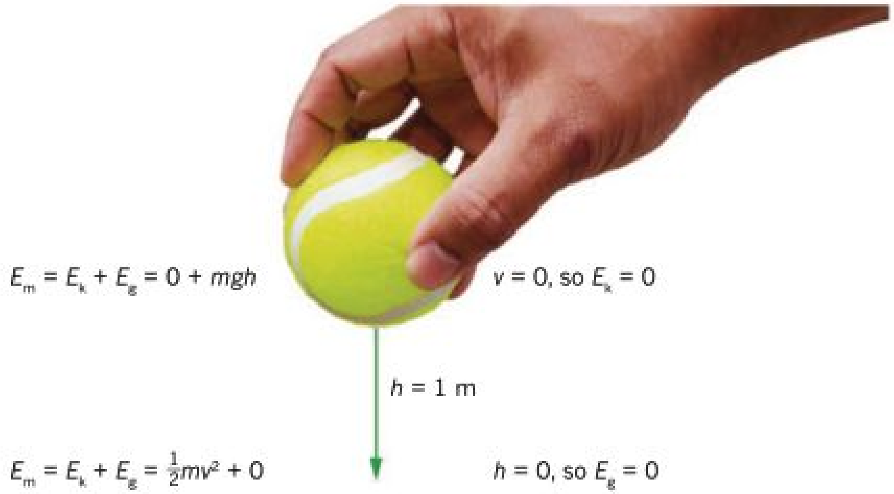


FIGURE 12.3.2 A falling tennis ball provides an example of conservation of mechanical energy.

At the instant the ball hits the ground, the total mechanical energy would comprise the gravitational potential energy available to it and the kinetic energy just prior to hitting the ground. The gravitational potential energy is 0 J because the ball is at ground level. To calculate kinetic energy, find the ball's velocity just before it hits the ground using one of the equations of motion:

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

PHYSICSFILE

Mechanical energy of a ball falling through the air

In reality, as a ball drops through the air, a very small amount of its energy is transformed into heat and sound and the ball won't quite reach a speed of 4.43 m s^{-1} before it hits the ground. This means that mechanical energy is not entirely conserved. However, this small effect can be considered negligible for many falling objects.

Principle of Conservation of Mechanical Energy

Given that, in a system of bodies, there are no other forms of energy except kinetic energy and potential energy, then the total mechanical energy of the system is constant.

Since $s = -1.00$ m, $a = -9.80$ m s^{-2} and $u = 0$ m s^{-1}:

$$v^2 = u^2 + 2as$$
$$v^2 = 0^2 + 2(-9.80 \times -1.00)$$
$$v = \sqrt{19.6}$$
$$= 4.43 \text{ m s}^{-1}$$

Therefore the kinetic energy of the tennis ball just before it hits the ground is:

$$E_\text{k} = \tfrac{1}{2}mv^2 = \tfrac{1}{2} \times 0.0600 \times 4.43^2 = 0.588 \text{ J}$$

Notice that at both the top and the bottom of the 1.00 m fall, the mechanical energy is the same. At the top:

$$E_\text{m} = E_\text{k} + E_\text{g} = 0 + 0.588 = 0.588 \text{ J}$$

At the bottom:

$$E_\text{m} = E_\text{k} + E_\text{g} = 0.588 + 0 = 0.588 \text{ J}$$

In fact, mechanical energy is constant throughout the drop. Consider the tennis ball when it has fallen halfway to the ground. At this point, $h = 0.500$ m and $v = 3.13$ m s^{-1}:

$$E_\text{m} = E_\text{k} + E_\text{g}$$
$$= (\tfrac{1}{2} \times 0.0600 \times 3.13^2) + (0.0600 \times 9.80 \times 0.500)$$
$$= 0.294 + 0.294$$
$$= 0.588 \text{ J}$$

Notice that, at this halfway point, the mechanical energy is evenly split between kinetic energy (0.294 J) and gravitational potential energy (0.294 J).

Throughout the drop, the mechanical energy has been conserved.

Worked example 12.3.1

MECHANICAL ENERGY OF A FALLING OBJECT

A basketball with a mass of 600 g is dropped from a height of 1.2 m. Calculate its kinetic energy at the instant before it hits the ground.

Thinking	Working
Since the ball is dropped, its initial kinetic energy is zero.	$(E_\text{k})_\text{initial} = 0$ J
Calculate the initial gravitational potential energy of the ball.	$(E_\text{g})_\text{initial} = mgh$ $= 0.600 \times 9.8 \times 1.2$ $= 7.1$ J
Calculate the initial mechanical energy.	$(E_\text{m})_\text{initial} = (E_\text{k})_\text{initial} + (E_\text{g})_\text{initial}$ $= 0 + 7.1$ $= 7.1$ J
At the instant before the ball hits the ground, its gravitational potential energy is zero.	$(E_\text{g})_\text{final} = 0$ J
Mechanical energy is conserved in this situation.	$(E_\text{m})_\text{initial} = (E_\text{m})_\text{final} = (E_\text{k})_\text{final} + (E_\text{g})_\text{final}$ $\therefore 7.1 = (E_\text{k})_\text{final} + 0$ $\therefore (E_\text{k})_\text{final} = 7.1$ J

Worked example: Try yourself 12.3.1

MECHANICAL ENERGY OF A FALLING OBJECT

A 6.8 kg bowling ball is dropped from a height of 0.75 m. Calculate its kinetic energy as it hits the ground.

USING MECHANICAL ENERGY TO CALCULATE VELOCITY

The speed of a falling object does not depend on its mass. This can be demonstrated using mechanical energy.

Consider an object with a mass, m, dropped from a height, h. At the moment it is dropped, its initial kinetic energy is zero. At the moment before it hits the ground, its final gravitational potential energy is zero. Therefore, using **conservation of mechanical energy**:

$$(E_m)_{\text{initial}} = (E_m)_{\text{final}}$$

$$(E_k)_{\text{initial}} + (E_g)_{\text{initial}} = (E_k)_{\text{final}} + (E_g)_{\text{final}}$$

$$\therefore 0 + mgh = \frac{1}{2}mv^2 + 0$$

$$\therefore mgh = \frac{1}{2}mv^2$$

$$\therefore gh = \frac{1}{2}v^2$$

$$\therefore v^2 = 2gh$$

$$\therefore v = \sqrt{2gh}$$

This formula can be used to find the velocity of a falling object as it hits the ground. Note that the formula does not contain the mass of the falling object, so if air resistance is negligible, any object with any mass will have the same final velocity when it is dropped from the same height.

EXTENSION

Deriving a formula for velocity from the equations of linear motion

The velocity of a falling object formula can also be derived from the equations of linear motion. Consider an object dropped from a height, h, with an initial speed of $u = 0$ m s^{-1}. Using the formula $v^2 = u^2 + 2as$:

$$\begin{aligned} v^2 &= u^2 + 2as \\ &= 0^2 + 2gh \\ &= 2gh \\ v &= \sqrt{2gh} \end{aligned}$$

This formula is equivalent to the result achieved using the conservation of mechanical energy.

Worked example 12.3.2

FINAL VELOCITY OF A FALLING OBJECT

A basketball with a mass of 600 g is dropped from a height of 1.2 m. Calculate the speed of the basketball at the instant before it hits the ground.

Thinking	Working
Recall the velocity of the falling object formula.	$v = \sqrt{2gh}$
Substitute the relevant values into the formula and solve.	$v = \sqrt{2 \times 9.8 \times 1.2} = 4.8$ m s^{-1}
Interpret the answer.	The basketball will be falling at 4.8 m s^{-1} just before it hits the ground.

Worked example: Try yourself 12.3.2

FINAL VELOCITY OF A FALLING OBJECT

A 6.8 kg bowling ball is dropped from a height of 0.75 m. Calculate the speed of the bowling ball just before it hits the ground.

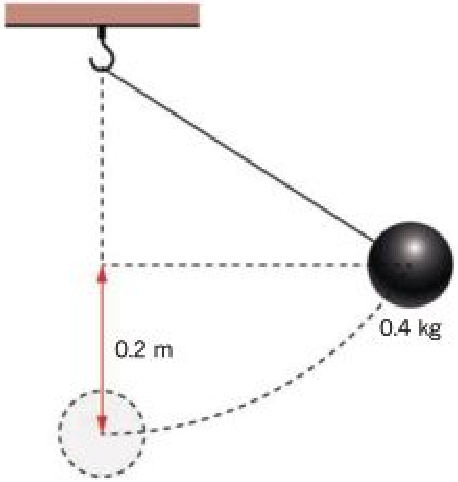


FIGURE 12.3.3 A falling pendulum provides an example of conservation of mechanical energy.

USING CONSERVATION OF MECHANICAL ENERGY IN COMPLEX SITUATIONS

The concept of mechanical energy allows physicists to determine outcomes in non-linear situations where the equations of linear motion cannot be used. For example, consider a pendulum with a bob of mass 400 g displaced from its mean position such that its height has increased by 20 cm, as shown in Figure 12.3.3.

Since a falling pendulum involves gravitational potential energy being converted into kinetic energy, the conservation of mechanical energy applies to this situation. Therefore, the formula developed earlier for the velocity of a falling object can be used to find the velocity of the pendulum bob at its lowest point.

$v = \sqrt{2gh} = \sqrt{2 \times 9.8 \times 0.2} = 2\ \text{m s}^{-1}$

The speed of the pendulum bob will be 2 m s^{-1} at its lowest point. However, unlike the falling tennis ball, the direction of the bob's motion will be horizontal instead of vertical at its lowest point. The equations of motion relate to linear motion and cannot be applied to this situation as the bob swings in a curved path.

Conservation of energy can also be used to analyse projectile motion – that is, when an object is thrown or fired into the air with some initial velocity. Since energy is not a vector, no vector analysis is required, even if the initial velocity is at an angle to the ground.

Worked example 12.3.3

USING MECHANICAL ENERGY TO ANALYSE PROJECTILE MOTION

A cricket ball (m = 140 g) is thrown upwards into the air at a speed of 15 m s^{-1}. Calculate the speed of the ball when it has reached a height of 8.0 m. Assume that the ball is thrown from a height of 1.5 m.

Thinking	Working
Recall the formula for mechanical energy.	$E_m = E_k + E_g = \frac{1}{2}mv^2 + mgh$
Substitute in the values for the ball as it is thrown.	$(E_m)_{\text{initial}} = \frac{1}{2}(0.14 \times 15^2) + (0.14 \times 9.8 \times 1.5)$ $= 18\ \text{J}$
Use conservation of mechanical energy to find an equation for the final speed.	$(E_m)_{\text{final}} = (E_k)_{\text{final}} + (E_g)_{\text{final}}$ $= \frac{1}{2}mv^2 + mgh$ $18 = \frac{1}{2}(0.14)v^2 + (0.14 \times 9.8 \times 8.0)$
Solve the equation algebraically to find the final speed.	$18 = 0.07v^2 + 11$ $7 = 0.07v^2$ $v^2 = \frac{7}{0.07}$ $v^2 = 100$ $v = \sqrt{100}$ $= 10\ \text{m s}^{-1}$
Interpret the answer.	The cricket ball will be moving at 10 m s^{-1} when it reaches a height of 8.0 m.

Worked example: Try yourself 12.3.3

USING MECHANICAL ENERGY TO ANALYSE PROJECTILE MOTION

An arrow with a mass of 35 g is fired into the air at 80 m s^{-1} from a height of 1.4 m. Calculate the speed of the arrow when it has reached a height of 30 m.

PHYSICS IN ACTION

Ballistics pendulum

The ballistics pendulum is an example of how the law of conservation of mechanical energy can be combined with an understanding of collisions to solve a practical problem. A ballistics pendulum is a device that can be used to measure the speed of a bullet fired from a gun or rifle. It consists of a block of wood hanging at a convenient height above the ground as shown in Figure 12.3.4.

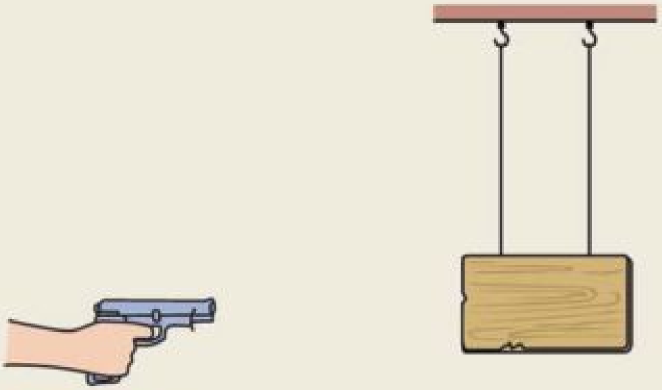


FIGURE 12.3.4 A ballistics pendulum combines an understanding of collisions and mechanical energy.

When a bullet is fired into the wooden block, an inelastic collision occurs. This means that much of the bullet's kinetic energy is converted into heat and sound and into changes made to the shape of the block. The conservation of mechanical energy does not apply for the impact of the bullet with the block.

However, the law of conservation of momentum still applies to the impact. This means that the block gains velocity from the bullet and it swings backwards and upwards as shown in Figure 12.3.5. By measuring the change in height of the block and the masses of the bullet and block, the initial speed of the bullet can be calculated. Note that conservation of mechanical energy does occur when the block swings backwards and upwards as no energy is converted into sound or heat during this part of its motion.

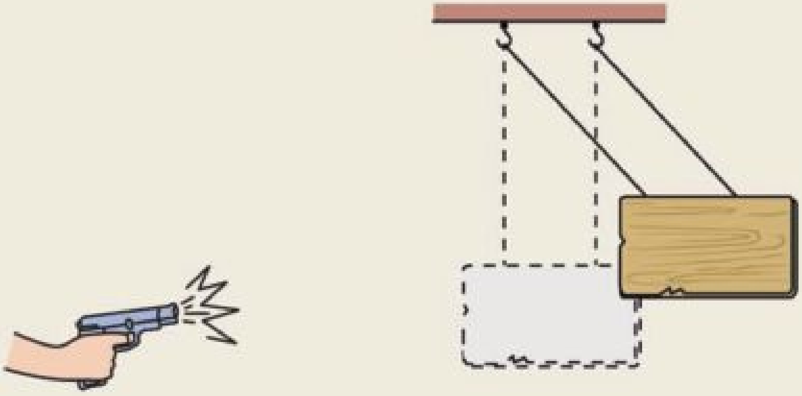


FIGURE 12.3.5 The change in height of a ballistics pendulum can be used to calculate the speed of the bullet fired into it.

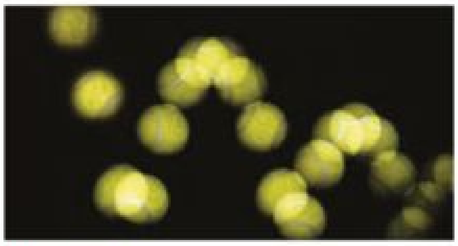


FIGURE 12.3.6 Mechanical energy is lost with each bounce of a tennis ball.

LOSS OF MECHANICAL ENERGY

Mechanical energy is not conserved in every situation. For example, when a tennis ball bounces a number of times, each bounce is lower than the one before it, as shown in Figure 12.3.6.

While mechanical energy is largely conserved as the ball moves through the air, a significant amount of kinetic energy is transformed into heat and sound when the ball compresses and decompresses as it bounces. This means that the ball does not have as much kinetic energy when it leaves the ground as it did when it landed. Therefore, the gravitational potential energy it can achieve on the second bounce will be less than the gravitational potential energy it had initially, and so the second bounce is lower.

EXTENSION

Elastic potential energy

A bouncing ball involves forms of energy other than kinetic and gravitational potential energy. When the ball hits the ground, its gravitational potential energy is converted into elastic potential energy as it compresses. As the ball expands back to its original shape, some of the elastic potential energy is converted back into kinetic energy and some of it is converted into heat and sound. The amount of energy that is converted into heat and sound depends on the type of ball.

If you want the ball to reach a greater height than its original height, then instead of dropping the ball you could add to its energy by throwing it downwards with some velocity. Consider Figure 12.3.7(a) where a tennis ball (m = 58 g) is thrown downwards at 4.0 m s^{-1} from a height of 1 m.

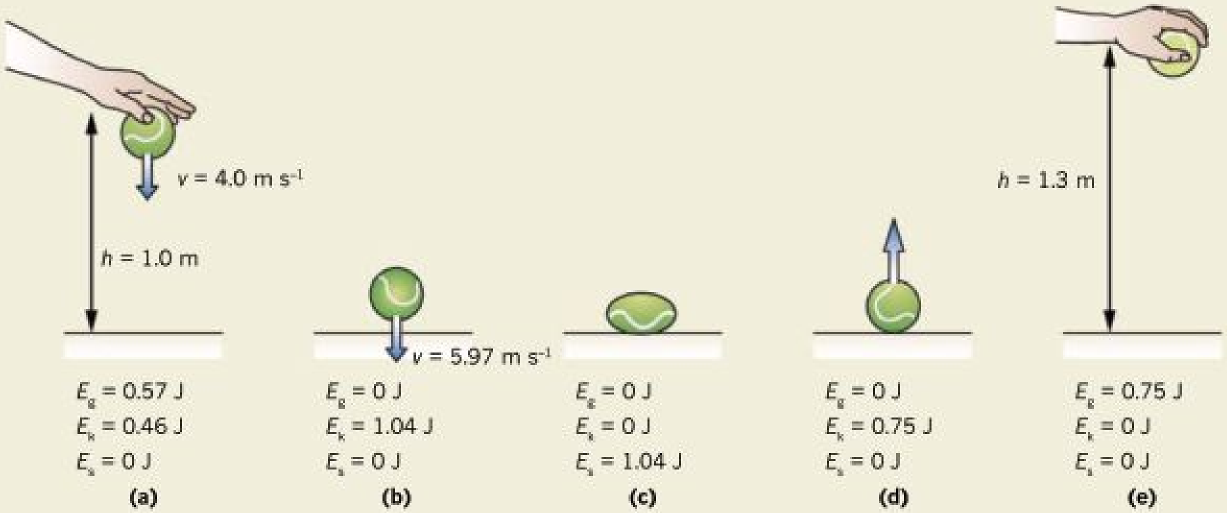


FIGURE 12.3.7 A tennis ball thrown downwards from a height.

Initially, the ball has 0.57 J of gravitational potential energy:

$mgh = 0.058 \times 9.8 \times 1.0 = 0.57$ J

and 0.46 J of kinetic energy:

$\frac{1}{2}mv^2 = \frac{1}{2}(0.058 \times 4^2) = 0.46$ J

By the time it reaches the ground, the gravitational potential energy has been transformed into kinetic energy, giving it a total of 1.04 J of kinetic energy (see Figure 12.3.7(b)). This is converted into elastic potential energy of 1.03 J (see Figure 12.3.7(c)).

If 0.28 J of energy are lost as heat and sound as the ball expands, then the ball will have just 0.75 J of kinetic energy when it leaves the ground (see Figure 12.3.7(d)). This means that it will rebound to a height of 1.3 m (see Figure 12.3.7(e)):

$h = \frac{E_g}{mg} = \frac{0.75}{0.058 \times 9.8} = 1.3$ m

Even though some energy has been 'lost' in the bounce, the initial kinetic energy of the ball means that it ends up slightly higher than where it started.

EFFICIENCY OF ENERGY TRANSFORMATIONS

In the real world, energy transformations are never perfect – there is always some energy 'lost'. Because of this, for a system to continue operating (doing work), it must be constantly provided with energy. The percentage of energy that is effectively transformed by a device is called the **efficiency** of that device. A device operating at 45% efficiency is converting 45% of its supplied energy into the useful new form. The other 55% is 'lost' or transferred to the surroundings, usually as heat and/or sound. It is not truly lost since energy cannot be created or destroyed; rather, the form it becomes (heat and sound) is not useful.

PHYSICSFILE

Coefficient of restitution

The 'bounce of the ball' is an important factor in many sports. Physicists describe the 'bounciness' of balls using a concept known as the coefficient of restitution (COR). COR depends on both the ball and the surface it is bouncing on. A tennis ball bouncing on grass has a different COR than one bouncing on clay. This is one reason why some tennis players prefer to play on some surfaces rather than others.

The efficiency of a transfer from one energy form to another is expressed as:

$$\text{Efficiency } (\eta) = \frac{\text{(useful energy transferred)}}{\text{(total energy supplied)}} \times 100\%$$

$$= \frac{\text{energy outrput}}{\text{energy input}} \times 100\%$$

Table 12.3.1 includes the approximate efficiencies of some common devices.

Device	Energy transfer	Efficiency (%)
electric motor	electric to kinetic	90
gas heater	chemical to thermal	75
incandescent light globe	electric to light	2
compact fluorescent light	electric to light	10
LED household light	electric to light	15
steam turbine	thermal to kinetic	45
coal-fired generator	chemical to electrical	30
high-efficiency solar cell	radiation to electrical	35
car engine	chemical to kinetic	25
open fireplace	chemical to thermal	15
human body	chemical to kinetic	25

TABLE 12.3.1 Efficiencies of some common devices.

Worked example 12.3.4

ENERGY EFFICIENCY

The energy input of a particular gas-fired power station is 1100 MJ. The electrical energy output is 300 MJ.

What is the efficiency of the power station in achieving this energy transfer?

Thinking	Working
Recall the equation for efficiency. Substitute the given values into the equation.	output = 300 MJ input = 1100 MJ $\text{efficiency } (\eta) = \frac{\text{energy output}}{\text{energy input}} \times 100\%$ $\text{efficiency } (\eta) = \frac{300\text{ MJ}}{1100\text{ MJ}} \times 100\%$
Solve the equation.	efficiency = 27%

Worked example: Try yourself 12.3.4

ENERGY EFFICIENCY

An electric kettle uses 23.3 kJ of electrical energy as it boils a quantity of water. The efficiency of the kettle is 18%.

How much electrical energy is expended in actually boiling the water?

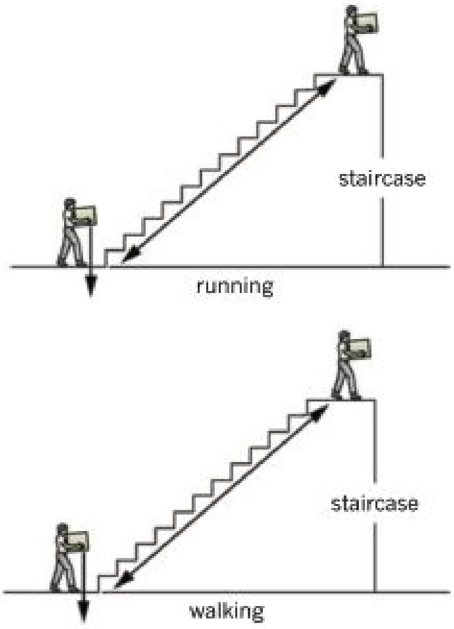


FIGURE 12.3.8 The runner and the walker both do the same amount of work, but the power output of the runner is higher than that of the walker.

PHYSICSFILE

Horsepower

James Watt was a Scottish inventor and engineer. He developed the concept of horsepower as a way to compare the output of steam engines with that of horses, which were the other major source of mechanical energy available at the time. Although the unit of one horsepower (1 hp) has had various definitions over time, the most commonly accepted value today is around 750 W. This is actually a significantly higher amount than an average horse can sustain over an extended period of time.

DEFINING POWER

Power is a measure of the rate at which work is done. Mathematically:

$$P = \frac{W}{\Delta t}$$

Recall that when work is done, energy is transferred or transformed. So the equation can also be written as:

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

where P is the power (in W)

E is the energy transferred or transformed (in J)

t is the time taken (in s).

For example, a person running up a set of stairs does exactly the same amount of work as if they had walked up the stairs (i.e. $W = mgh$). However, the rate of energy change is faster for running up the stairs. Therefore, the runner is applying more power than the walker (see Figure 12.3.8).

Unit of power

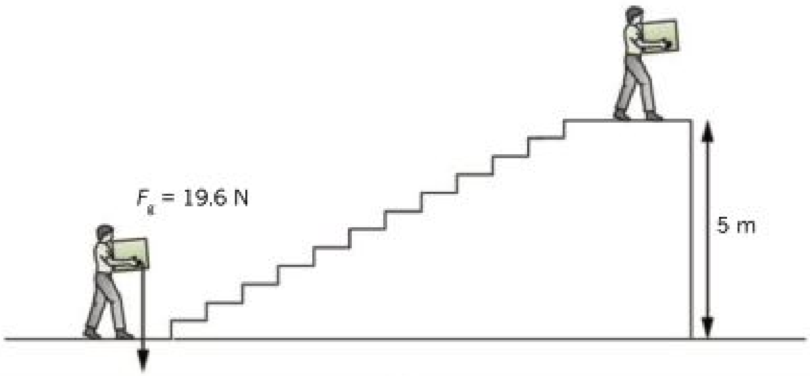
The unit of power is named after the Scottish engineer James Watt, who is most famous for inventing the steam engine. A watt (W) is defined as a rate of work of one joule per second; in other words:

$$1\text{ W} = \frac{1\text{ J}}{1\text{ s}} = 1\text{ J s}^{-1}$$

Worked example 12.3.5

CALCULATING POWER

Calculate the power required to carry a box with a mass of 2 kg up a 5 m staircase in 20 s. (Use $g = 9.8$ m s^{-2}.)



Thinking	Working
Calculate the force applied.	$F_g = mg$ $= 2 \times 9.8$ $= 19.6$ N
Calculate the work done.	$W = Fs$ $= 19.6 \times 5$ $= 98$ J
Recall the formula for power.	$P = \frac{W}{\Delta t}$
Substitute the appropriate values into the formula.	$P = \frac{98}{20}$
Solve.	$P = 4.9$ W

Worked example: Try yourself 12.3.5

CALCULATING POWER

Calculate the power used by a weightlifter to lift a barbell which has a total mass of 50 kg from the floor to a height of 2.0 m above the ground in 1.4 s. (Use $g = 9.8\ \text{m s}^{-2}$.)

POWER, FORCE AND AVERAGE SPEED

In many everyday situations, a force is applied to an object to keep it moving at a constant speed, for example, pushing a wardrobe across a carpeted floor or driving a car at a constant speed. In these situations, the power being applied can be calculated directly from the force applied and the speed of the object.

Since $P = \frac{W}{\Delta t}$ and $W = Fs$, then:

$$P = \frac{Fs}{\Delta t} = F \times \frac{s}{\Delta t}$$

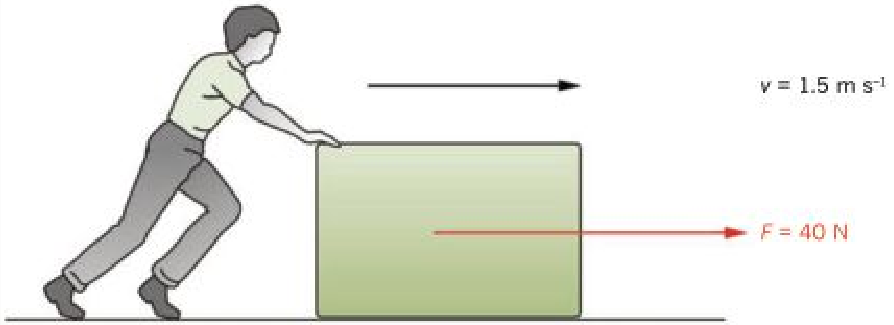
Since $\frac{s}{\Delta t}$ is the definition of average speed, the power equation can be written as:

$$P = Fv_{av}$$

Worked example 12.3.6

FORCE–VELOCITY FORMULATION OF POWER

A person pushes a heavy box along the ground at an average speed of 1.5 m s^{-1} by applying a force of 40 N. What amount of power does the person exert on the box?



Thinking	Working
Recall the force–velocity formulation of the power equation.	$P = Fv_{av}$
Substitute the appropriate values into the formula.	$P = 40 \times 1.5$
Solve.	$P = 60$ W

Worked example: Try yourself 12.3.6

FORCE–VELOCITY FORMULATION OF POWER

Calculate the power required to keep a car moving at an average speed of 22 m s^{-1} if the force of friction (including air resistance) is 1200 N. Give your answer correct to three significant figures.

12.3 Review

SUMMARY

- Mechanical energy is the sum of the potential and kinetic energies of an object.
- Mechanical energy is conserved in a falling object.
- Conservation of mechanical energy can be used to predict outcomes in a range of situations involving gravity and motion.
- The final velocity of an object falling from height h can be found using the equation $v = \sqrt{2gh}$.
- When a ball bounces, some mechanical energy is transformed into heat and sound.
- Power is a measure of the rate at which work is done: $P = \frac{W}{\Delta t} = \frac{\Delta E}{\Delta t}$.
- The power required to keep an object moving at a constant speed can be calculated from the product of the force applied and its average speed: $P = Fv_{av}$.
- The efficiency of an energy transfer from one form to another is:
  - Efficiency $(\eta) = \frac{\text{(useful energy transferred)}}{\text{(total energy supplied)}} \times 100\%$ $= \frac{\text{energy output}}{\text{energy input}} \times 100\%$

KEY QUESTIONS

1 A piano with a mass of 180 kg is pushed off the roof of a five-storey apartment block. The piano falls 3 m for each storey (i.e. a total of 15 m).

a Calculate the piano's kinetic energy as it hits the ground.

b Calculate the piano's kinetic energy as it passes the windows on the second floor, having fallen 10 m.

2 A tennis ball is dropped from the roof of a five-storey apartment block. The tennis ball falls 3 m for each storey (i.e. a total of 15 m).

a Calculate the tennis ball's speed as it hits the ground.

b Calculate the tennis ball's speed as it passes the windows on the second floor, having fallen 10 m.

3 A branch falls from a tree and hits the ground with a speed of 5.4 m s^{-1}. From what height did the branch fall?

4 A javelin with a mass of 800 g is thrown at an angle of inclination of 40°. It is released at a height of 1.45 m with a speed of 28.5 m s^{-1}.

a Calculate the javelin's initial mechanical energy.

b Calculate the speed of the javelin as it hits the ground.

5 A coal-fired generator has an efficiency of approximately 30%. If 2000 J of energy is supplied to the generator, how much is converted into electrical energy?

6 A rubber ball is dropped from a height of 1.5 m and loses 20% of its mechanical energy as it hits the ground. To what height will it rebound?

7 A 1610 kg car accelerates from zero to 100 km h^{-1} in 5.50 s. Calculate its average power output over this time.

8 A locomotive engine applies a force of 4 kN to keep a train moving at 20 m s^{-1}. Calculate the power output of the engine.

9 A 1700 kg car's engine uses 40 kW of power to maintain a constant speed of 80 km h^{-1}. Calculate the force being applied by its engine.

10 The motor of a crane has a maximum power output of 25 kW. At what average speed could it lift a concrete slab with a mass of 500 kg?

Chapter review

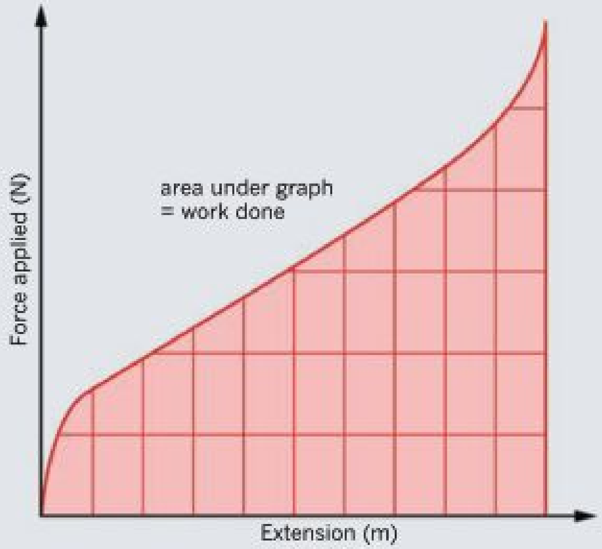
12

KEY TERMS

conservation of mechanical energy
efficiency
elastic
elastic potential energy
gravitational potential energy
Hooke's law
mechanical energy
power
spring constant

1 A car drives at a constant speed for 80 m. In order to overcome friction, its engine applies a force of 2000 N. Calculate the work done by the engine.

2 Estimate the total work shown in the following force–displacement graph, given that each grid square corresponds to 1 J.



3 A crane lifts a 200 kg load from the ground to a height of 30 m. What is the work done by the crane?

4 A person walks up a flight of 12 stairs. Each step is 240 mm long and 165 mm high. If the person has a mass of 60 kg and $g = 9.8\ \text{m s}^{-2}$, what is the total amount of work done against gravity?

5 If 4000 J is used to lift a 50.0 kg object with a constant velocity, what is the theoretical maximum height to which the object can be raised?

6 A pram is pushed by the handle, which is at an angle of 35° to the horizontal (the direction of motion). If 1200 J of work is done pushing the pram 20 m, with what force was the pram pushed?

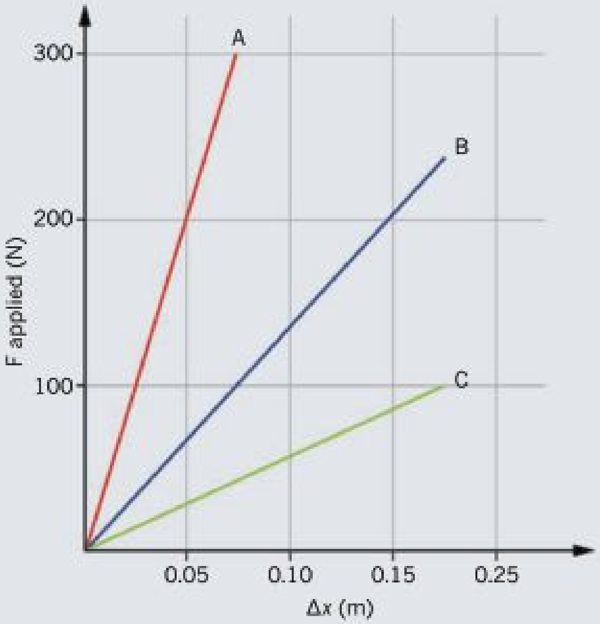
7 A cricket ball with a mass of 160 g is bowled with a speed of 150 km h^{-1}. What is the kinetic energy of the cricket ball?

8 If a 1200 kg car has kinetic energy of 70 kJ, what is its speed?

9 The speed of an object is doubled. By how much does its kinetic energy increase?

10 A plumber (mass 88 kg) digs a ditch 40 cm deep. By how much does the plumber's gravitational potential energy change when he steps from the ground down into the ditch?

11 The force–extension graph for three different springs is shown below.



Calculate the approximate spring constant for each spring.

12 A football with a mass of 0.43 kg is kicked off the ground with a speed of 16 m s^{-1}. How fast will it be going when it hits the crossbar, which is 2.44 m above the ground?

13 A bullet of mass 10 g strikes a ballistics pendulum of mass 1.5 kg with speed v and becomes embedded in the pendulum. When the pendulum swings back, its height increases by 15 cm. For the questions that follow, assume that the initial gravitational potential energy of the pendulum was zero.
- **a** What was the gravitational potential energy of the pendulum at the top of its swing?
- **b** What was the kinetic energy of the pendulum when the bullet first embedded in it?
- **c** What was the speed of the pendulum when it started to swing?

Chapter review *continued*

14 A crane can lift a load of 5 tonnes vertically through a distance of 20 m in 5 s. What is the power of the crane approximately?

15 A red Mini Cooper with a mass of 650 kg can accelerate from 0 to 100 km h^{-1} in 7.2 s. What is the average power output of the car over this time?

16 If the engine of a 1400 kg car uses 25 kW to maintain an average speed of 17 m s^{-1}, how much friction is acting on the car?

17 At the start of a 100 m race, a runner with a mass of 60 kg accelerates from a standing start to 8.0 m s^{-1} in a distance of 20 m.

a Calculate the work done by the runner's legs.

b Calculate the average force that the runner's legs apply over this distance.

18 When moving around on the Moon, astronauts find it easier to use a series of small jumps rather than to walk. If an astronaut (with a mass of 120 kg including equipment) jumps to a height of 10 cm on the Moon, where the gravitational field strength is 1.6 m s^{-2}, by roughly how much does his potential energy increase?

19 The efficiency of an appliance is known to be 80%. What energy was supplied if the output was 1250 J?

UNIT 2 • Area of Study 1

REVIEW QUESTIONS

How can motion be described and explained?

1 An aeroplane flies a distance of 300 km due north, then changes course and travels 400 km due east. What is the distance travelled and the final displacement of the aeroplane?

- **A** distance = 700 km, displacement = 500 km north east
- **B** distance = 700 km, displacement = 700 km north east
- **C** distance = 700 km, displacement = 500 km north 53.1° east
- **D** distance = 700 km, displacement = 500 km north 36.9° east

The following information relates to questions 2 and 3. A car accelerates in a straight line at a rate of 5.5 m s^{-2} from rest.

2 What distance has the car travelled at the end of three seconds?

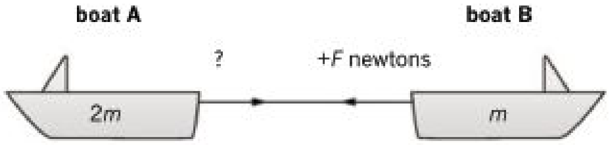
- **A** 8.25 m
- **B** 11 m
- **C** 16.5 m
- **D** 24.75 m

3 What distance does the car travel in the third second of its motion?

- **A** 8.25 m
- **B** 13.75 m
- **C** 19.25 m
- **D** 24.75 m

4 Two boats are tied together with an inextensible rope as shown in the diagram below. Boat A has twice the mass of boat B.

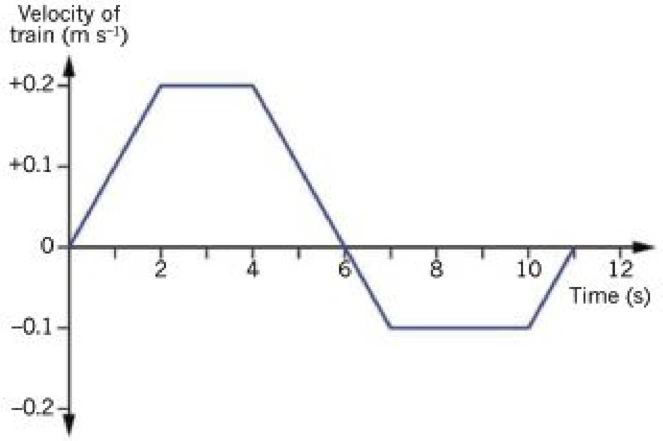
Boat A exerts a force of +*F* newtons on boat B.



What is the magnitude and direction of the force on boat A from boat B?

- **A** $-\frac{1}{2}F$ N
- **B** +*F* N
- **C** –*F* N
- **D** –2*F* N

The following information relates to questions 5 and 6. A graph depicting the velocity of a small toy train versus time is shown below. The train is moving on a straight section of track, and is initially moving in the easterly direction.



5 What distance does the train travel in the first 6 seconds of its motion?

- **A** 0 m
- **B** 0.4 m
- **C** 0.8 m
- **D** 1.2 m

6 What is the displacement of the train after the first 11 seconds of its motion?

- **A** 0 m
- **B** 0.4 m east
- **C** 0.8 m east
- **D** 1.2 m east

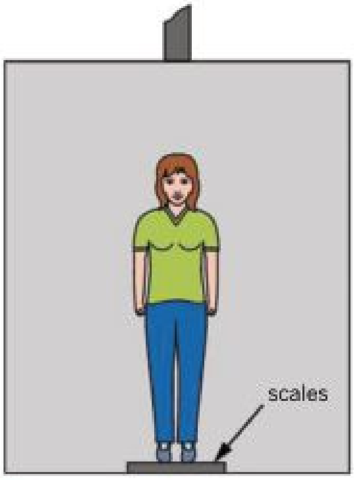
7 A ball dropped from rest from a height *h* hits the ground with a velocity *v*. The ball is then released from a height of 2*h*. With what velocity would the ball now strike the ground?

- **a** $\frac{1}{2}v$
- **b** $\sqrt{2}v$
- **c** 2*v*
- **d** 4*v*

8 A ball is dropped, falls vertically and strikes the ground with a velocity of +5 m s^{-1}. It rebounds, and leaves the ground with a velocity of –3 m s^{-1}. What is the change in velocity that the ball experiences?

- **A** –8 m s^{-1}
- **B** +8 m s^{-1}
- **C** –2 m s^{-1}
- **D** +2 m s^{-1}

The following information relates to questions 9 and 10. A 60 kg student stands on a set of digital scales in an elevator, as shown in the diagram below.



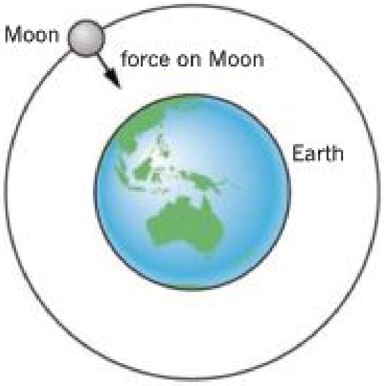
9 In which of the following situation/s will the digital scales show a reading of 60 kg?

A when the lift is travelling upwards at constant velocity
B when the lift is travelling downwards at constant velocity
C when the lift is stationary
D all of the above

10 The lift travels upwards with an acceleration of a. Taking up as positive (+), which of the following is the correct expression for the normal force (N) on the student? m = mass of the student, g = acceleration due to gravity

A $N = mg + ma$
B $N = mg - ma$
C $N = ma - mg$
D $-N = mg + ma$

11 The orbit of the Moon around Earth can be modelled as circular motion in which the Moon can be considered to orbit at a fixed height (radius) above the surface of Earth and at a constant speed. Earth exerts a gravitational force on the Moon that acts at right angles to the velocity of the Moon as shown in the diagram below.



Which of the following statements about the Moon is correct?

A The Moon experiences no change in kinetic energy during an orbit.
B The Moon experiences no change in gravitational potential energy during an orbit.
C Earth's gravitational force does no work on the Moon.
D All of the above statements are correct.

12 Two dynamics carts are being used in the physics laboratory to study momentum. The two carts have the same mass, m kg, and are travelling towards each other with a speed v m s^{-1}, as shown in the diagram. Take the right-hand direction to be positive.

What is the total momentum of this system?

A 0 kg m s^{-1}
B $+mv$ kg m s^{-1}
C $+2mv$ kg m s^{-1}
D $-2mv$ kg m s^{-1}

13 A single spring has a spring constant of 10 N m^{-1}. What mass must be suspended from the spring to cause the spring to stretch (extend) by 20 cm? Choose the closest answer.

A 2 g
B 20 g
C 200 g
D 2000 g

14 An Olympic archery competitor tests a bow by firing an arrow of mass 25 g vertically into the air. The arrow leaves the bow with an initial vertical velocity of 100 m s^{-1}. The acceleration due to gravity may be taken as $g = 9.8$ m s^{-2} and the effects of air resistance can be ignored.

a At what time will the arrow reach its maximum height?
b What is the maximum vertical distance that this arrow reaches?
c What is the acceleration of the arrow when it reaches its maximum height?

15 A hiker travels 8 km in a northerly direction from his campsite and then travels a further 7 km in a north easterly direction. What is his final displacement?

16 If the journey takes a total of 7 hours, calculate the average speed of the hiker in question 15 in m s^{-1}.

17 Two students drop a lead weight from a tower and time its fall. How far does the weight travel during the second second, compared with the first second?

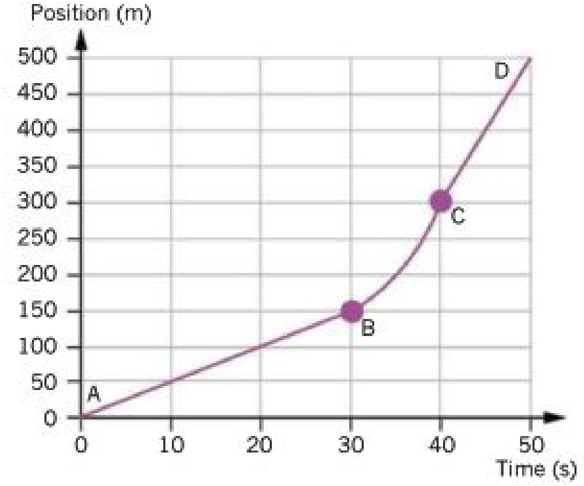
18 A car with good brakes, but smooth tyres, has a maximum retardation of 4.0 m s^{-2} on a wet road. The driver has a reaction time of 0.50 s. The car is travelling at 72 km h^{-1} when the driver sees a danger and reacts by braking.

- **a** How far does the car travel during the reaction time?
- **b** Assuming maximum retardation, calculate the braking time.
- **c** Determine the total distance travelled by the car from the time the driver realises the danger to the time the car finally stops.

19 Helen and Emily conduct the following experiment from a skyscraper. Helen drops a platinum sphere from a vertical height of 122 m while at exactly the same time Emily throws a lead sphere with an initial downwards vertical velocity of 10.0 m s^{-1} from a vertical height of 140 m. Assume g = 9.80 m s^{-2} and ignore friction.

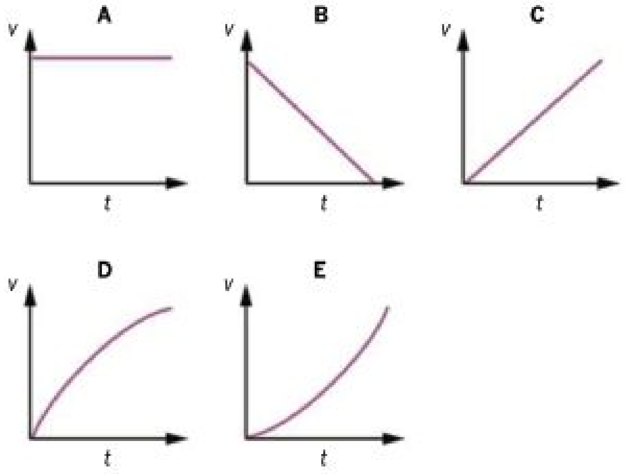
- **a** Determine the time taken by the platinum sphere to strike the ground.
- **b** Calculate the time taken by the lead sphere to strike the ground.
- **c** Determine the average velocity of each sphere over their respective distances.

20 The following position versus time graph depicts the motion of a cyclist travelling east along a straight road from points A to D.



- **a** Describe the motion of the cyclist in terms of speed.
- **b** What was the velocity of the cyclist for the first 30 s?
- **c** What was the velocity of the cyclist for the final 10 s?
- **d** Calculate the average velocity between points B and C.
- **e** Calculate the average acceleration between points B and C.
- **f** Calculate the average speed between points A and D.

The following velocity versus time graphs refer to questions 21 and 22.



21 **a** Which of the graphs show acceleration?

- **b** If the graphs are taken to depict the same time interval, which one represents the greatest displacement.

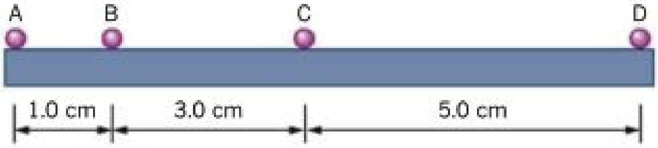
22 Which graph could describe the following motion?

- **a** a ball being dropped from a height with air resistance neglected
- **b** a skydiver jumping out of a plane allowing for air resistance
- **c** an ice skater gliding on the ice with negligible friction

The following information applies to questions 23 to 25.

$$\frac{x}{t} = \frac{3.0 \times 10^{-2}\ \text{m}}{0.10\ \text{s}}$$

During a physics experiment a student sets a multiflash timer at a frequency of 10 Hz. A nickel marble is rolled across a horizontal table. The diagram shows the position of the marble for the first four flashes: A, B, C and D. Assume that when flash A occurred t = 0, at which time the marble was at rest.



23 Determine the average speed of the marble for these distance intervals:

- **a** A to B
- **b** B to C
- **c** C to D

24 Determine the instantaneous speeds of the marble for these times:

- **a** t = 0.05 s
- **b** t = 0.15 s
- **c** t = 0.25 s

25 Describe the motion of the marble.

26 A 100 kg man is standing at rest on the ground. Use $g = 9.8\ \text{m s}^{-2}$.

- **a** Name the forces acting on the man, using the description F_{AB} to describe the force on A by B.
- **b** Indicate the relative magnitudes of these forces.
- **c** Describe the reaction forces that form action–reaction pairs with the forces on the man.

27 Three wooden blocks each of weight 100 N are stacked one above the other on a table. Block C is on the table, block B is in the middle and block A is on top. Use the symbols T for table, E for Earth and then A, B and C for the blocks to answer the following questions.

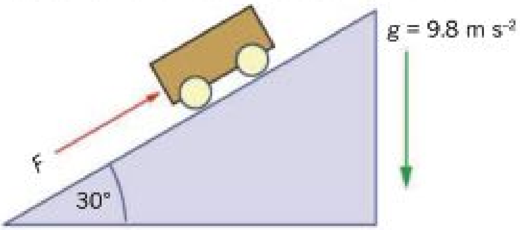
- **a** Name the forces acting on block A and give the magnitude and direction of each.
- **b** Name the forces acting on block B and give the magnitude and direction of each.
- **c** Likewise, name the forces acting on C and give the magnitude and direction of each.
- **d** A child quickly knocks the bottom block C out from under the other two. Calculate the net force on each of blocks A and B and hence describe their motion.

28 A tow truck, pulling a car of mass 1000 kg along a straight road, causes the velocity of the car to increase from 5.00 m s^{-1} west to 10.0 m s^{-1} west in a distance of 100 m. A constant frictional force of 200 N acts on the car.

- **a** Calculate the acceleration of the car.
- **b** What is the resultant force acting on the car during this 100 m?
- **c** Calculate the force exerted on the car by the tow truck.
- **d** What force does the car exert on the tow truck?

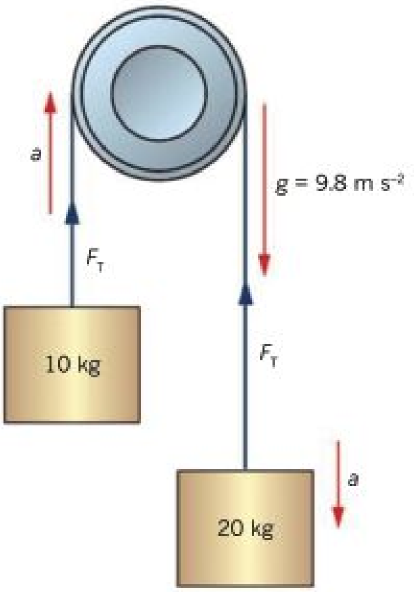
29 A car that is initially at rest begins to roll down a steep road that makes an angle of 11.3° with the horizontal. Ignoring friction, determine the speed of the car in km h^{-1} after it has travelled a distance of 100 m. (Assume $g = 9.8\ \text{m s}^{-2}$.)

30 A 100 kg trolley is being pushed up a rough 30° incline by a constant force F. The frictional force F_f between the incline and the trolley is 110 N.



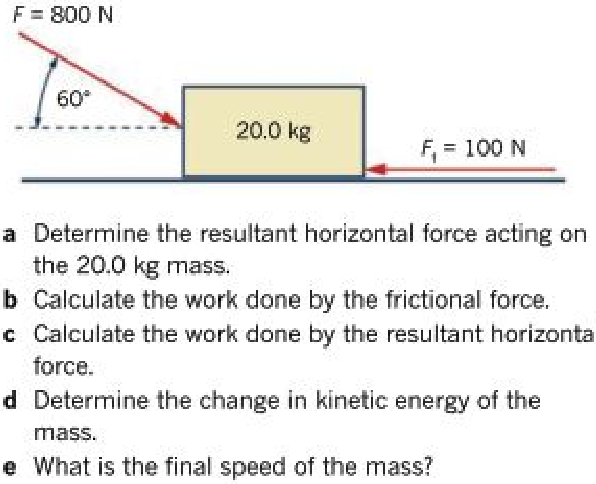
- **a** Determine the value of F that will move the trolley up the incline at a constant velocity of 5.0 m s^{-1}.
- **b** Determine the value of F that will accelerate the trolley up the incline at 2.0 m s^{-2}.
- **c** Calculate the acceleration of the trolley when $F = 1000$ N.

31 Two masses, 10 kg and 20 kg, are attached via a steel cable to a frictionless pulley, as shown in the following diagram.



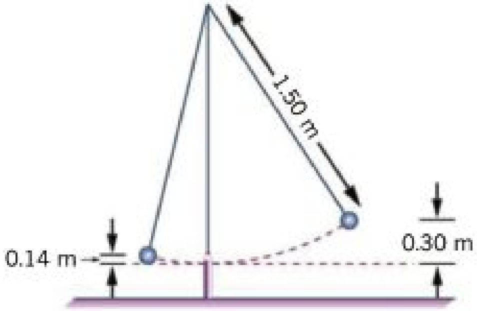
- **a** Determine the acceleration for each mass.
- **b** What is the magnitude of the tension in the cable?

32 An 800 N force is applied as shown to a 20.0 kg mass, initially at rest on a horizontal surface. During its subsequent motion the mass encounters a constant frictional force of 100 N while moving along a horizontal distance of 10 m.



- **a** Determine the resultant horizontal force acting on the 20.0 kg mass.
- **b** Calculate the work done by the frictional force.
- **c** Calculate the work done by the resultant horizontal force.
- **d** Determine the change in kinetic energy of the mass.
- **e** What is the final speed of the mass?

The following information applies to questions 33 to 35. A mass of 0.40 kg hangs from a string 1.5 m long. The string is kept taut and the mass is drawn aside a vertical distance of 0.30 m, as shown in the diagram below. A pencil is fixed in a clamp so that when the mass is released it will swing down and break the pencil. The mass swings on but now only moves through a vertical distance of 0.14 m. (Assume $g = 9.8\ \text{m s}^{-2}$.)



33 Calculate the velocity of the mass the instant before it strikes the pencil.

34 Calculate the work required to break the pencil.

35 Account for the loss in energy.

36 Two equal forces are applied to a rigid object. Explain how the object can be in translational equilibrium but not in rotational equilibrium.

37 Calculate the torque on a refrigerator door of width 900 mm if a force of 500 N is applied to open it.

38 Three children are playing on a see-saw. Susan has a mass of 35 kg and is sitting 3.0 m from the pivot point. Her younger brother James has mass 25 kg and is sitting 3.5 m from the pivot on the same side as Susan. Where should their older brother Thomas (mass 42 kg) position himself in order to balance his two younger siblings?

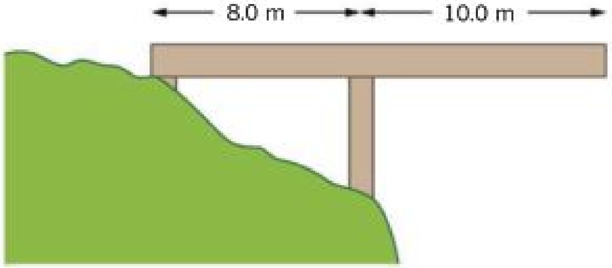
The following information applies to questions 39 and 40. A wheelbarrow measures 1.70 m from the handles to the wheel axle. The wheelbarrow itself has mass 20 kg and the centre of mass is assumed to be 50 cm from the axle.

39 If a load of 85 kg is placed in the wheelbarrow at a distance of 70 cm from the axle, what is the minimum force required to lift the wheelbarrow?

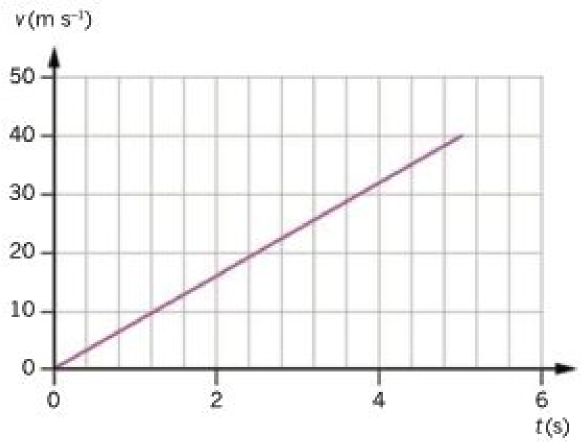
40 A gardener pushes this wheelbarrow a distance of 20 m over level ground, applying a force of 600 N at an angle to the wheelbarrow handles.

- **a** Calculate the angle at which the force is applied so that there is just sufficient upwards force to counter the weight of the load and wheelbarrow.
- **b** Calculate the work done in pushing the wheelbarrow.
- **c** Explain why no work is done by the vertical component of the force.
- **d** Calculate the power of the gardener over his 20 m journey if he moves at a constant speed of $2.0\ \text{m s}^{-1}$.

41 Calculate the upwards forces applied by the ground and the supporting column on the cantilever beam shown below given that the mass of the beam is 300 kg.

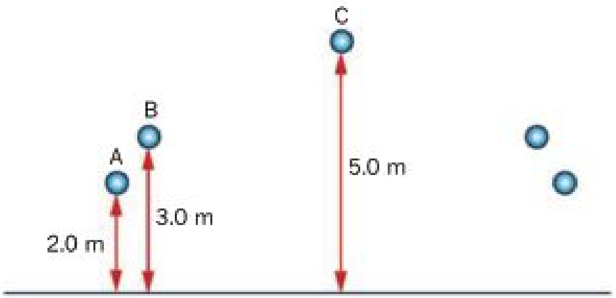


42 The figure shows the velocity–time graph for a car of mass 2000 kg. The engine of the car is providing a constant driving force. During the 5.0 s interval the car encounters a constant frictional force of 400 N. At $t = 5.0$ s, $v = 40.0\ \text{m s}^{-1}$.



- **a** How much kinetic energy (in MJ) does the car have at $t = 5.0$ s?
- **b** What is the resultant force acting on the car?
- **c** What force is provided by the car's engine during the 5.0 s interval?
- **d** How much work is done on the car during the 5.0 s interval?
- **e** Determine the power output of the car's engine during the 5.0 s interval.
- **f** How much heat energy is produced due to friction during the 5.0 s interval?
- **g** Calculate the efficiency with which the energy provided by the engine is transformed into kinetic energy.

43 The following diagram shows the trajectory of a 2.0 kg shotput recorded by a physics student during a practical investigation. The sphere is projected at a vertical height of 2.0 m above the ground with initial speed $v = 10\ \text{m s}^{-1}$. The maximum vertical height of the shotput is 5.0 m. (Ignore friction and assume $g = 9.8\ \text{N kg}^{-1}$.)



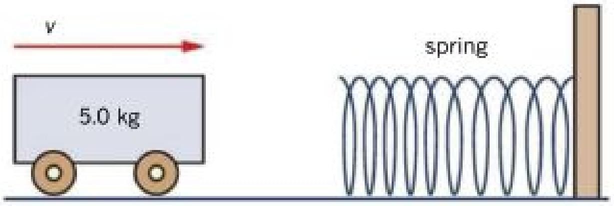
a What is the total energy of the shotput just after it is released at point A?

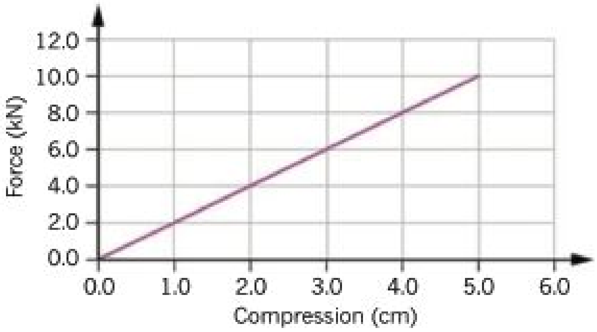
b What is the kinetic energy of the shotput at point B?

c What is the minimum speed of the shotput during its flight?

d What is the total energy of the shotput at point C?

44 A 5.0 kg trolley approaches a spring that is fixed to a wall. During the collision, the spring undergoes a compression Δx and the trolley is momentarily brought to rest, before bouncing back at $10\ \text{m s}^{-1}$. The force–compression graph for the spring is shown below. (Ignore friction.)





a Calculate the elastic potential energy stored in the spring when its compression is equal to 2.0 cm.

b What is the elastic potential energy stored in the spring when the trolley momentarily comes to rest?

c At what compression will the trolley come to rest?

d Explain why the trolley starts moving again.

e What property of a spring accounts for the situation described above?

The following information applies to questions 45 and 46. A child of mass 34 kg drops from a height of 3.50 m above the surface of a trampoline. When the child lands on the trampoline, it stretches so that she is 50 cm below the initial trampoline position. (Use $g = 9.8\ \text{m s}^{-1}$).

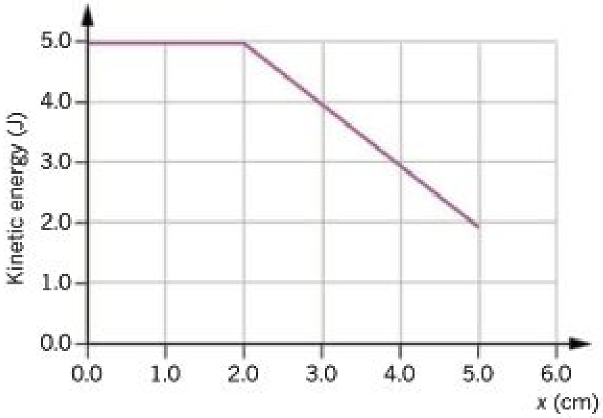
45 **a** What is the spring constant of the trampoline?

b The child rebounds to a height of 0.75 m below her original position. Calculate the efficiency of the energy transfer of the trampoline.

c At what point does the child have maximum kinetic energy?

46 Describe the energy transformations that occur as the child drops, bounces and rebounds as described in the question above.

47 A nickel cube of mass 200 g is sliding across a horizontal surface. One section of the surface is frictionless while the other is rough. The graph shows the kinetic energy, E_k, of the cube versus distance, x cm, along the surface.



a Which section of the surface is rough? Justify your answer.

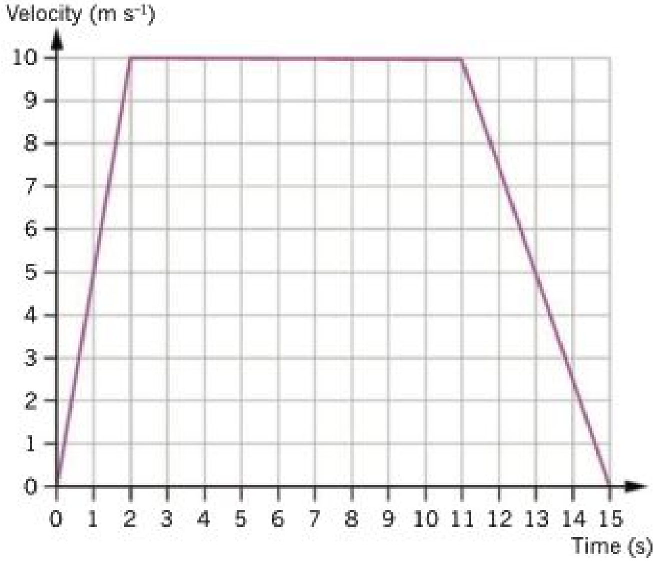
b Determine the speed of the cube during the first 2.0 cm.

c How much kinetic energy is lost by the cube between $x = 2.0$ cm and $x = 5.0$ cm?

d What has happened to the kinetic energy that has been lost by the cube?

e Calculate the value of the average frictional force acting on the cube as it travels over the rough surface.

The following information applies to questions 48 and 49. The diagram is an idealised velocity–time graph for the motion of an Olympic sprinter.



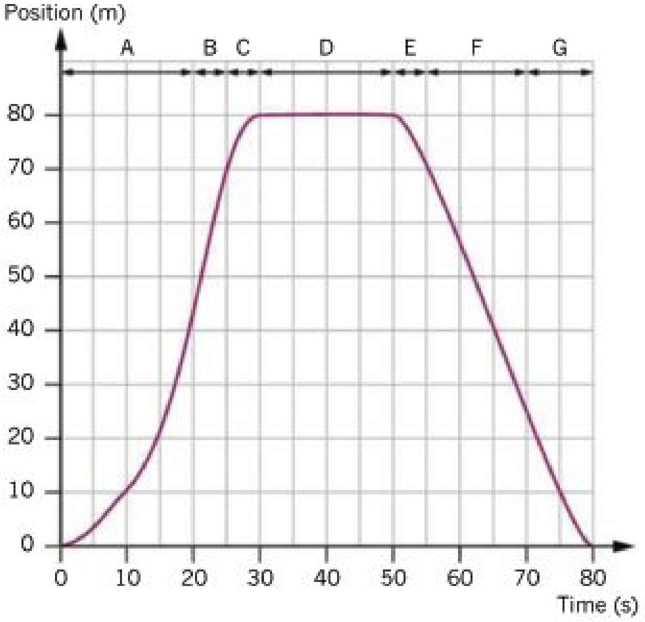
48 What distance was this race?

49 Determine the average speed of the sprinter:

a while she is racing to the finish line

b for the total time that she is moving.

The following information applies to questions 50 to 52. The diagram gives the position–time graph of the motion of a boy on a bicycle. The boy initially travels in a northerly direction.



50 During which of the section(s) (A–G) is the boy:

a travelling north?

b stationary?

c travelling south?

d speeding up?

e slowing down?

51 For the boy's 80 s ride, calculate:

a the total distance covered

b the average speed.

52 Determine the average velocity of the boy:

a when t = 10 s

b during section B

c when t = 60 s.

53 A small car is found to slow down from 90 km h^{-1} to 60 km h^{-1} in 12 s when the engine is switched off and the car is allowed to coast on level ground. The car has a mass of 830 kg.

a What is the car's deceleration (in m s^{-2}) during the 12 s interval?

b What was the average braking force acting on the car during the time interval?

c Determine the distance that the car travels during the 12 s interval.

d Explain what happens to all the initial kinetic energy of the car.

54 A swimmer propels herself through the water with her arms. Explain her motion in terms of Newton's laws.

55 A spaceship with a mass of 20 tonnes (2.0×10^4 kg) is launched from the surface of Earth, where g has a value of 9.8 N kg^{-1} downwards, to land on the Moon, where g is 1.6 N kg^{-1} downwards. What is the weight of the spaceship when it is on Earth and when it is on the Moon?

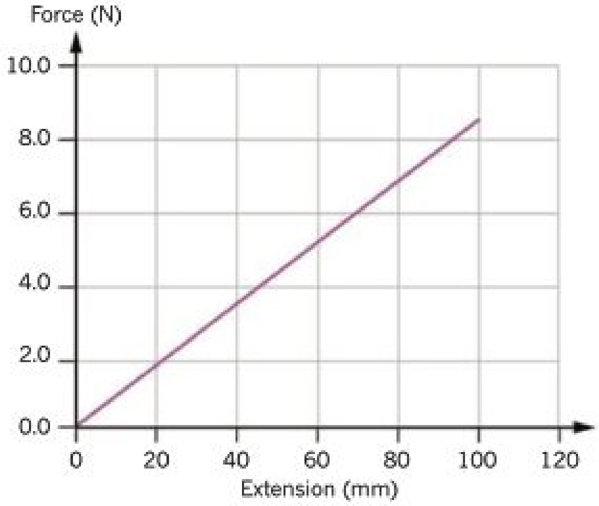
56 **a** Describe the motion of your chair when you stand up and push it back from your desk.

b How would the chair behave if it were on castors?

c Explain how your answers to parts a and b are illustrations of Newton's laws.

57 An engine pulls a line of rail cars along a flat track with a steady force, but instead of accelerating, the whole train travels at a constant velocity. How can this be consistent with Newton's first and second laws of motion?

58 Students were conducting an experiment to investigate the behaviour of springs. Increasing masses (m) were hung from a vertically suspended spring and the resulting force–extension graph was plotted as shown.



- **a** Estimate the value of the spring constant.
- **b** Use your answer to part a to calculate the elastic potential energy stored in the spring when the extension is 100 cm.

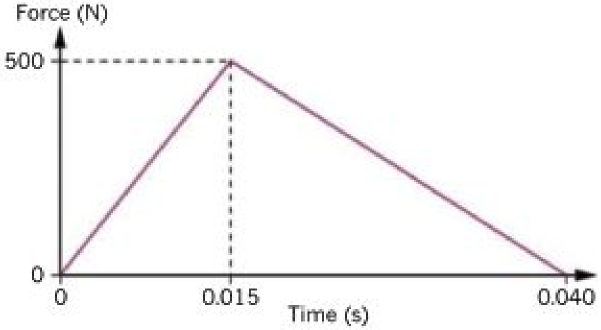
59 A football player of mass 86 kg travelling at 7.5 m s^{-1} collides with a goalpost.

- **a** Calculate the impulse of the post on the football player.
- **b** Is momentum conserved in this collision? Explain.
- **c** If he hits his head, explain, using Newton's laws, why he may sustain a concussion.

60 A goods train wagon of mass 4.0×10^4 kg travelling at 3.0 m s^{-1} in a shunting yard couples with a stationary wagon of mass 1.5×10^4 and the two wagons move on at a reduced speed.

- **a** Calculate the speed of the coupled wagons.
- **b** Check to see if kinetic energy is conserved in this collision.
- **c** Explain your finding.

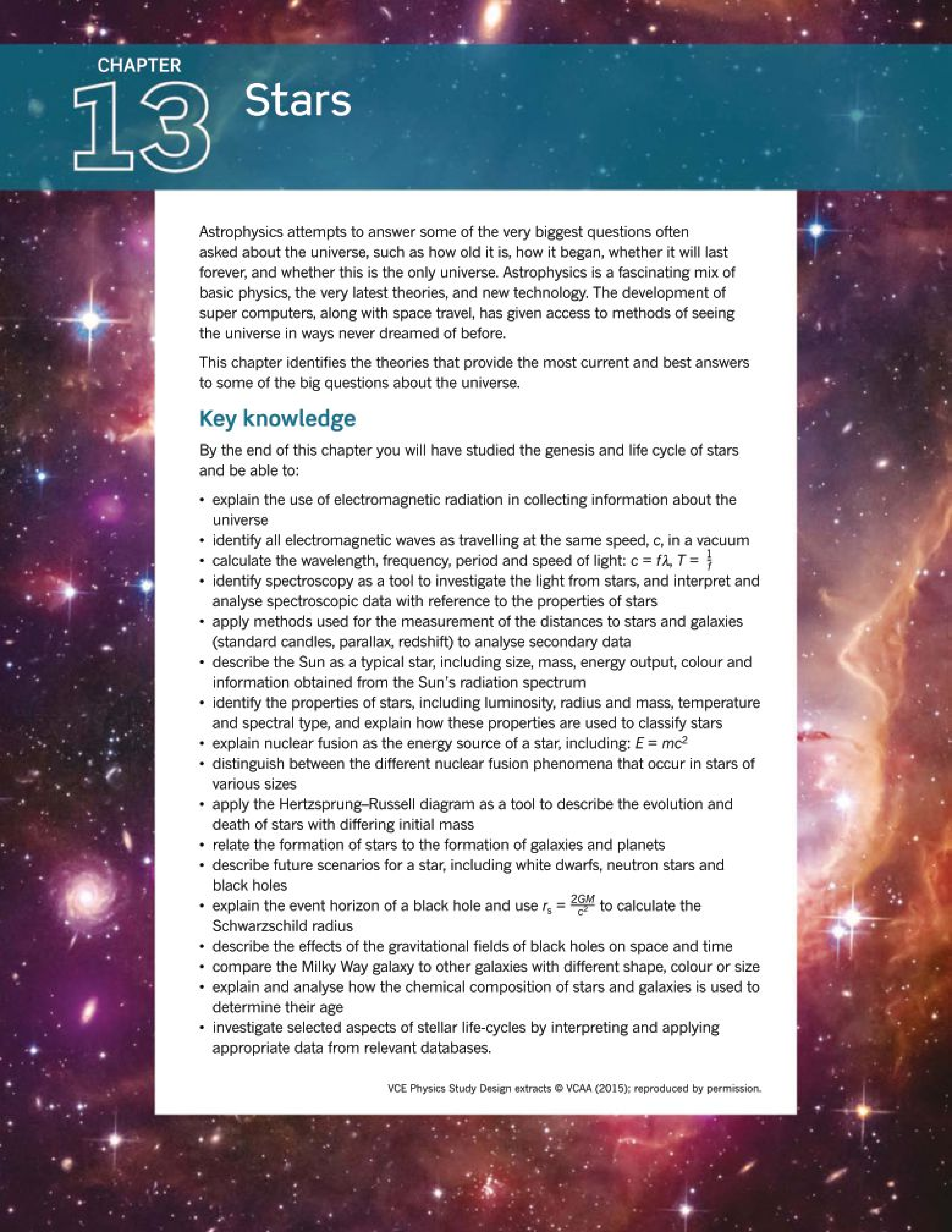
61 Jordy is playing softball and hits a ball with her softball bat. The force–time graph for this interaction is shown below. The ball has a mass of 170 g. Assume that the bat and ball form an isolated system during the interaction.



- **a** Determine the impulse experienced by the ball.
- **b** Calculate the change in momentum of the bat.
- **c** Determine the magnitude of the change in velocity of the ball.

62 A naval cannon with mass 1.08×10^5 kg fires projectiles of mass 5.5×10^2 kg with a muzzle speed of 8.0×10^2 m s^{-1}. The barrel of the cannon is 20 m long, and it can be assumed that the propellant acts on the projectile for the time that it is in the barrel.

- **a** Calculate the magnitude of the average acceleration of the projectile down the barrel.
- **b** Using Newton's second law, calculate the average force exerted by the propellant as the projectile travels down the barrel.
- **c** Calculate the momentum of the projectile as it leaves the barrel.
- **d** Calculate the recoil velocity of the gun.
- **e** Calculate the average force of the propellant from the change in momentum of the projectile.
- **f** Calculate the average work done by the propellant on the projectile and compare this with the kinetic energy gained by the projectile.



CHAPTER 13

Stars

Astrophysics attempts to answer some of the very biggest questions often asked about the universe, such as how old it is, how it began, whether it will last forever, and whether this is the only universe. Astrophysics is a fascinating mix of basic physics, the very latest theories, and new technology. The development of super computers, along with space travel, has given access to methods of seeing the universe in ways never dreamed of before.

This chapter identifies the theories that provide the most current and best answers to some of the big questions about the universe.

Key knowledge

By the end of this chapter you will have studied the genesis and life cycle of stars and be able to:

- explain the use of electromagnetic radiation in collecting information about the universe
- identify all electromagnetic waves as travelling at the same speed, c, in a vacuum
- calculate the wavelength, frequency, period and speed of light: $c = f\lambda$, $T = \frac{1}{f}$
- identify spectroscopy as a tool to investigate the light from stars, and interpret and analyse spectroscopic data with reference to the properties of stars
- apply methods used for the measurement of the distances to stars and galaxies (standard candles, parallax, redshift) to analyse secondary data
- describe the Sun as a typical star, including size, mass, energy output, colour and information obtained from the Sun's radiation spectrum
- identify the properties of stars, including luminosity, radius and mass, temperature and spectral type, and explain how these properties are used to classify stars
- explain nuclear fusion as the energy source of a star, including: $E = mc^2$
- distinguish between the different nuclear fusion phenomena that occur in stars of various sizes
- apply the Hertzsprung–Russell diagram as a tool to describe the evolution and death of stars with differing initial mass
- relate the formation of stars to the formation of galaxies and planets
- describe future scenarios for a star, including white dwarfs, neutron stars and black holes
- explain the event horizon of a black hole and use $r_s = \frac{2GM}{c^2}$ to calculate the Schwarzschild radius
- describe the effects of the gravitational fields of black holes on space and time
- compare the Milky Way galaxy to other galaxies with different shape, colour or size
- explain and analyse how the chemical composition of stars and galaxies is used to determine their age
- investigate selected aspects of stellar life-cycles by interpreting and applying appropriate data from relevant databases.

13.1 Astronomical measurements

Early astronomers thought the stars all lay in the 'realm of the gods', just beyond the planets that were circling the Earth, which was itself stationary at the centre of the universe. It was Galileo who first found evidence that not all objects circled the Earth. With the new invention of the telescope he was able to see that Jupiter had moons which orbited it. So he and some other astronomers began to support a theory first put forwards by Copernicus that the Earth circled the Sun and that the stars must be much further away than the planets.

By Newton's time it was realised that the stars were possibly other 'suns' and therefore must be huge distances away or they would appear much brighter. Many people, however, refused to believe the stars could be suns. For thousands of years it had been believed that they were 'heavenly' objects quite unlike anything in the solar system, and located just beyond the orbit of Saturn. The sort of distances that Galileo and Newton were suggesting were simply incomprehensible.

Scientists now know that the universe is incredibly big and our solar system is just a very small speck in our galaxy, which is just one of billions of galaxies. Scientists believe there are as many galaxies in the universe as there are stars in the Milky Way galaxy. The universe is not fixed. Throughout the universe, new stars are forming (Figure 13.1.1) and galaxies are evolving.

FIGURE 13.1.1 The Omega Nebula 'star factory' is a huge cloud of dust and molecular gases in which massive stars are forming.

This chapter extends some of the ideas that were introduced in Chapter 5 'The origins of everything', which talked about the origins of the universe.

STARLIGHT AND THE ELECTROMAGNETIC SPECTRUM

'Light' is a simple way of referring to the complete electromagnetic spectrum that you would know from earlier chapters. The electromagnetic spectrum itself is much more than the visible section that can be detected by the human eye. The electromagnetic spectrum also includes the radio and infrared waves that are longer than visible light as well as the ultraviolet and X-rays that are shorter. The complete electromagnetic spectrum is key to physicists understanding the workings of the universe.

The only way that scientists can learn more about stars is by studying the light they emit. Light reveals details about the properties of stars—there will be more on this in Section 13.2 'Classifying stars'. Light also provides the methods, direct and indirect, by which astrophysicists are able to determine the distances between planets in our solar system and out to the stars and galaxies beyond.

SPEED AND FREQUENCY OF LIGHT

Observing a spectrum of colours, such as a rainbow, is nothing new for most people. However, what causes light to disperse (break up) into its component colours or frequencies is useful to know in order to understand some of the techniques used for determining distances in space.

The speed of electromagnetic radiation (light) of all frequencies is the same in a vacuum. The speed of light is denoted c, indicating that this speed, in a vacuum, is believed to be a constant throughout the universe. The speed of light is well known and can even be determined with reasonable precision in the standard physics classroom.

ℹ The speed of light, c, is constant in a vacuum and is approximately equal to 3.0×10^8 m s^{-1}.

In Einstein's famous equation $E = mc^2$, the speed of light acts as the constant linking mass and energy. In other words, c in this equation refers to the speed of light while E and m are energy and mass, respectively. This equation is explained in more detail on page 229.

While the speed of light is constant in a vacuum, when light travels from one medium to another its speed will change, and the speed will be different depending on the optical density of the medium and the frequency of the light. If light enters the medium at an angle, due to its change in speed, the light will also bend. You will have seen evidence of this phenomenon if you have ever seen a spoon or straw in a glass of water (see Figure 13.1.2). The spoon or straw will appear bent. The bending of light as it travels from one medium to another and changes speed is referred to as **refraction**.

How the speed of light changes can be explained by conservation of momentum and conservation of energy, but what is important for now is that there is a change and that the change also effects the wavelength of the light. The change in wavelength is proportional to the change in speed, and hence the light is separated into its component range of wavelengths. The frequency of the light itself doesn't change.

In other words, the speed of electromagnetic radiation for a particular medium is proportional to the wavelength:

$c \propto \lambda$

where λ is the wavelength in metres (m)

c is the speed of light in m s^{-1}.

As the frequency is constant, it is the constant of proportionality, so:

FIGURE 13.1.2 The refraction of light as it enters the water in the glass.

$c = f\lambda$

where c is the speed of light (3.0×10^8 m s^{-1})

f is the frequency of the radiation or the number of wavelengths per second passing a given point in hertz (Hz)

λ is the wavelength of the light in metres (m).

Since frequency is the number of wavelengths per second, then the time in seconds for one complete wavelength to pass a given point is referred to as the period, T.

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

where T is the period for one wavelength in seconds (s)

f is the frequency of light in hertz (Hz).

EXTENSION

The wave equation

$c = f\lambda$ is a particular form of the general wave equation: $v = f\lambda$.

The wave equation applies to any form of energy transfer that exhibits wave or wave-like behaviour. In this form of the equation, v is the velocity of the wave and if the wave form is light, then v is equal to c. Waves and the wave behaviour of light are studied in Unit 4, when this equation will be more extensively developed and applied.

Worked example 13.1.1

CALCULATING THE FREQUENCY AND PERIOD OF LIGHT

The speed of light in a vacuum is approximately 3.0×10^8 m s^{-1}. The average wavelength of visible light from the Sun is 483 nm.

a What is the corresponding frequency for this wavelength?

Thinking	Working
Determine what is known and unknown from the question and convert quantities to SI units.	$c = 3.0 \times 10^8$ m s^{-1} $\lambda = 483$ nm $= 483 \times 10^{-9}$ m $f = ?$
Rearrange the wave equation for light to make the unknown the subject.	$c = f\lambda$ $\frac{c}{\lambda} = f$ $f = \frac{c}{\lambda}$
Substitute the known values and calculate *f*. Round the answer to two significant figures to match the least-accurate figure supplied.	$f = \frac{3 \times 10^8}{483 \times 10^{-9}}$ $= 6.2 \times 10^{14}$ Hz

b What is the period of the light wave, i.e. the time for one complete wave to pass a given point?

Thinking	Working
Refer to the relationship between period and frequency.	$T = \frac{1}{f}$
Substitute the value for the frequency and solve for *T*, taking care to use SI units correctly.	$T = \frac{1}{f}$ $= \frac{1}{6.2 \times 10^{14}}$ $= 1.6 \times 10^{-15}$ s

Worked example: Try yourself 13.1.1

CALCULATING THE FREQUENCY AND PERIOD OF LIGHT

The speed of light in a vacuum is approximately 3.0×10^8 m s^{-1}. The wavelength of red light is 650 nm.

a What is the corresponding frequency for this wavelength?

b What is the period of the red light wave, i.e. the time for one complete wave to pass a given point?

Worked example 13.1.2

CALCULATING THE WAVELENGTH

The frequency of far-infrared light is approximately 3×10^{11} Hz. What is the wavelength of this light?

Thinking	Working
Write the wave equation in terms of λ.	$c = f\lambda$ $\lambda = \frac{c}{f}$
Substitute the values for the frequency and speed of light. Solve for λ.	$\lambda = \frac{3 \times 10^8}{3 \times 10^{11}}$ $= 0.001$ m or 1 mm

Worked example: Try yourself 13.1.2

CALCULATING THE WAVELENGTH

The frequency of a particular radio wave is 3×10^8 Hz. What is the wavelength of this light?

SPECTROMETERS AND SPECTROSCOPY

Using a glass prism, it is possible to separate light into its component colours. In order for scientists to be able to analyse the light from stars, more sophisticated forms of the same basic tool must be used to break that light down into its component wavelengths.

An alternative to the glass prism is a **diffraction grating**. While the image from a diffraction grating is not as bright as that from a prism, it has a significant advantage in that it is basically flat and can be used along with other tools for the direct measurement of the wavelength of the source across the whole of the electromagnetic spectrum.

Diffraction gratings are typically optically flat, highly transparent glass etched with hundreds or even thousands of precision-drawn lines. The unetched portions between the lines **diffract** the light passing through them, which forms an **interference pattern** separating the light into its component wavelengths (see Figure 13.1.3). You may have seen a similar 'rainbow' effect when light reflects off the finely etched surface of a media disc such as a CD, DVD or Blu-ray.

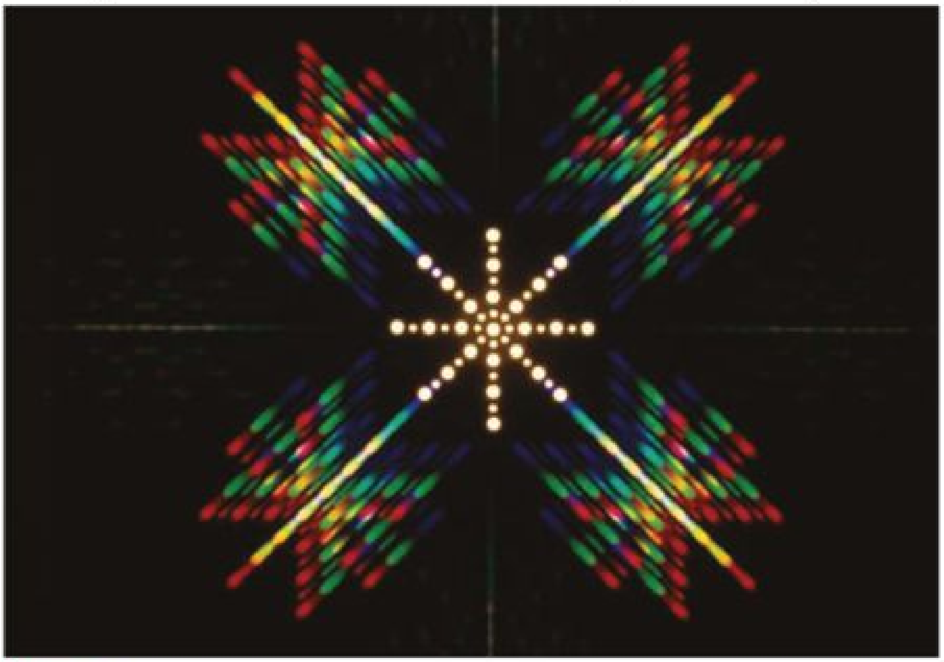


FIGURE 13.1.3 The diffraction pattern of a white light source produced by placing a diffraction grating over a camera lens. The diffraction pattern reveals the wavelengths present in the white light.

A **spectroscope** is a simple device combining a diffraction grating with a viewing telescope, which allows direct viewing of the spectrum of light from a source such as a distant star. When a graduated scale is added to the spectroscope, the specific wavelengths in the spectrum can be determined. This is known as a spectrometer and it has become key to analysing the spectra from distant stars. The branch of science investigating the spectra produced when matter interacts with or emits electromagnetic radiation is referred to as spectroscopy.

Modern spectrometers use electronic means to record the specific wavelengths of the incident radiation, but they function on the same basic principle. They are used by astronomers to investigate the light from stars, and in the next section 'Classifying stars', you will see how spectroscopy has become fundamental to investigating the properties of stars.

FIGURE 13.1.4 A model of the spectroscope invented by William Crooke in the nineteenth century.

REDSHIFT

FIGURE 13.1.5 (a) Gustav Robert Kirchhoff, 1824–87; (b) Robert Wilhelm Bunsen, 1811–99.

You may be familiar with the name 'Kirchhoff' from work on electrical circuits. Gustav Robert Kirchhoff (Figure 13.1.5(a)), a German physicist, is well known for his work on formulating theorems for determining the distribution of current in networks of electrical conductors. Kirchhoff also worked alongside Robert Wilhelm Bunsen (Figure 13.1.5(b)), of Bunsen burner fame, in developing a complete system for spectral analysis in 1859–60. Their work established spectroscopy as a means by which the chemical properties of stars could be determined by comparing the stars' spectra with those of known elements here on Earth. (Refer to Section 5.1 to refresh your understanding of emission spectra.)

William Huggins (1824–1910), a British astronomer, applied Kirchhoff and Bunsen's work on spectrum analysis to the astronomical observations he was making. His first results were presented in 1863. Huggins went further in attempting to photograph stellar spectra to record spectra for later analysis. He adapted new photographic techniques that were becoming available at the time to his work on spectroscopic analysis. His long-exposure photographs revealed objects that had previously been invisible through telescopes alone. While his main interest was in analysing the spectral composition of the light, he also noticed that the spectra of light from more distant objects was 'shifted' towards the red end of the spectrum. Huggins realised that this was due to the Doppler effect that also lengthens the wavelength of sounds as they move away from us. It was evidence that distant stars were moving away from Earth, as shown in Figure 13.1.6.

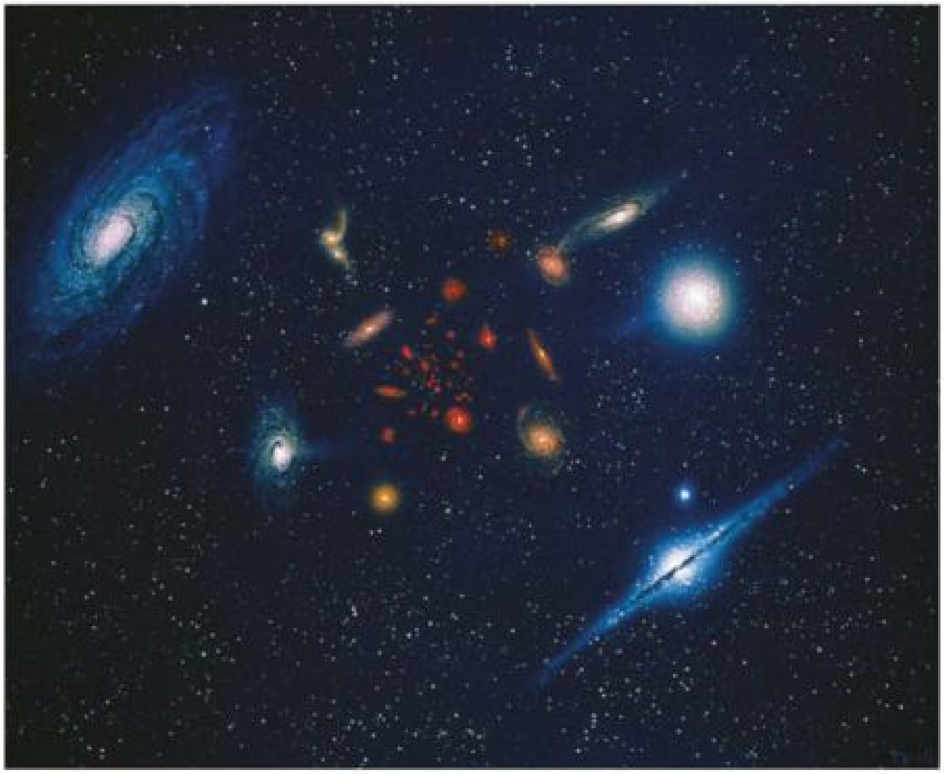


FIGURE 13.1.6 This image of a group of galaxies has been colour-coded to show how fast they are moving away from Earth. Those coloured red are moving away fastest. They will not necessarily appear red to the naked eye, it is just that their spectra have been shifted closer to the red end of the electromagnetic spectrum.

In 1929, Edward Hubble used measurements of this redshift to show that the galaxies moving away from Earth the fastest are also the most distant, providing evidence that the universe is expanding.

You will recall reading about Hubble and redshift previously in Section 5.1 on page 172. This section explains in detail how Hubble came to this conclusion and how it can be applied to determining the velocity at which a galaxy is moving away from Earth depending upon the distance it is away.

ASTRONOMICAL DISTANCES

Stellar parallax

The only way the astronomers who followed Galileo could measure the distance to the stars was to look for parallax movement in their observations taken as the Earth moved around the Sun.

As the Earth moves around its orbit, a closer star should appear to move in relation to those further away. It's a little like holding up a pen and looking at it while moving your head from side to side: the pen seems to move relative to the background.

Various eighteenth century astronomers attempted to measure the annual parallax of some bright stars that they thought were close to Earth. In 1729 James Bradley announced that if there was any parallax in the positions of stars, it was less than 1 second of arc (also written as 1" of arc). He could not find examples of parallax greater than this. He calculated that the stars he observed must be at least 400 000 times further away from us than the distance to the Sun from Earth (that is, 400 000 AU). Along with Newton's calculation of the distance to Sirius, at one million AU from Earth, this meant that any attempt to measure the distances to stars seemed futile and so there were few attempts in the remainder of the eighteenth century.

PHYSICSFILE

Seconds of arc

Standard astronomical units for distance are discussed in Chapter 5 'The origins of everything', in Table 5.1.1 on page 169.

The unit of 'seconds of arc' is in more common use in navigation, mapping and surveying as an alternative to radians, but you might also see it written as arc seconds (arcsec). You may also have come across it in mathematics.

A circle consists of 360°; each degree is broken down into 60 minutes (60') and each minute into 60 seconds (60"). One second of arc is therefore $\frac{1}{60} \times \frac{1}{60} = \frac{1}{3600}$ of a degree.

By the beginning of the 1800s instrument makers had developed special telescopes capable of measuring angles of only fractions of a second of arc, and so attempts were renewed to look for nearby stars showing parallax. By this time it had been realised that some stars move, relative to the background stars, at a steady rate. This was not due to parallax—the star was actually moving very slowly through the sky (proper motion). For example, the star 61 Cygni had been observed to move 5" of arc every year. At that rate it would take over 700 years to move 1°.

The early nineteenth century astronomers realised that stars with a large proper motion were more likely to be closer to Earth than those with none. It was also likely that brighter stars were closer to Earth.

The development of better telescopes had led to the discovery that many stars were actually **binary stars** (see Figure 13.1.7). That is, they appeared to be two stars in orbit around each other. The further apart the two stars appeared to be, the more likely that the stars would be closer to the Earth.

This gave three criteria for assessing how close a star was to Earth: significant proper motion, brightness and distance between binary stars. A systematic examination was made of stars that fulfilled at least two of these three criteria.

In 1835, German-born Wilhelm Struve looked at Vega, a bright single star, which had a large proper motion. He was able to determine that Vega showed a parallax angle of just one-eighth of a second of arc. That measurement suggested that Vega was 1.6 million AU from the Earth.

Soon other 'close' stars were found. In 1838 Thomas Henderson found that the Alpha Centauri system (shown in Figure 13.1.8), a bright binary system with a high proper motion, had a parallax of just under 1". A third faint star in the Centauri constellation called Proxima Centauri has a parallax of 0.77", making it the closest star to our Sun at a distance of 270 000 AU.

FIGURE 13.1.7 The Circinus X–1 binary stars. They are located around 26 000 light-years from Earth, in the Milky Way galaxy.

FIGURE 13.1.8 A coloured optical image of Alpha Centauri. Alpha Centauri A and B are a double star system and one of the nearest to our own system. They are about 20" apart.

This method of measuring the distance to stars is known as **stellar parallax**. It leads to a unit for the distance of stars called the parsec. The parsec (pc) is the distance to a star that shows a parallax angle of one second of arc. A star which has a parallax of only 0.5" will be twice as far away as one with 1" of parallax, and so on. The distance in parsec is therefore the reciprocal of the parallax angle (as explained in the Chapter 5 'The origins of everything').

$$d = \frac{1}{p}$$

where d is the distance in parsec

p is the parallax angle in seconds.

Worked example 13.1.3

FINDING DISTANCE FROM PARALLAX ANGLE

Alpha Centauri A and B are about 20" apart.

a Convert this parallax angle into parsec.	
Thinking	**Working**
Recall the rule for converting from parallax angle to parsec.	$d = \frac{1}{p}$
Apply the rule to find the distance.	$d = \frac{1}{20}$ $= 0.05$ pc

b Given that 1 pc is approximately 3×10^{16} m, how far apart are Alpha Centauri A and B, in metres?	
Thinking	**Working**
Use the relationship 1 pc = 3×10^{16} m	0.05 pc = $0.05 \times 3 \times 10^{16}$ m $= 1.5 \times 10^{15}$ m

Worked example: Try yourself 13.1.3

FINDING DISTANCE FROM PARALLAX ANGLE

Proxima Centauri is about 0.77" from Earth.

a Convert this parallax angle into parsec.

b Given that 1 pc is approximately 3×10^{16} m, how far from Earth is Proxima Centauri, in metres?

A limitation of parallax measurements is the distortion, or 'shimmer', of the star's image caused by the Earth's atmosphere. For this reason, in 1989, the European Space Agency launched a satellite called Hipparcos. Its purpose was to measure the stellar parallax of as many stars as possible from above the atmosphere. The name is an acronym for High Precision Parallax Collecting Satellite and honours the ancient Greek astronomer Hipparchus, who was one of the first to map the stars systematically. The satellite Hipparcos was able to measure the parallax of over 100 000 stars to an accuracy of 0.001", in other words out to nearly 1000 pc. This may seem a vast distance, but given that Earth's galaxy, the Milky Way, is about 50 000 pc wide and that Earth is about 8500 pc from its centre, a lot of stars are still 'out of range' for stellar parallax to be useful.

THE STANDARD CANDLE APPROACH TO DISTANCE

In the twentieth century, while parallax was used to measure the distance of many thousands of stars, most stars were too far away to use this form of measurement. As a result, astronomers looked for indirect means of determining the distance to a star. The most obvious was to find a way of determining the **absolute brightness** or **intrinsic brightness** (*L*), and to compare it to the **apparent brightness** (*b*) of a star as seen from Earth. The absolute brightness of a star is its actual brightness, and the apparent brightness is how bright it looks from a distance. This is called the **standard candle** approach to determining distance. If you knew that a certain star had the same absolute brightness (*L*) as our Sun, it would be possible to calculate how much further away it was by comparing its apparent brightness (*b*) to the apparent brightness ($b_{\odot}$) of the Sun. (The subscript $_{\odot}$ is used when referring to or comparing with the value for the Sun.) The problem is how to determine absolute brightness (*L*).

Apparent brightness

Originally, astronomers measured the apparent brightness (*b*) of stars using a scale that was created by Greek astronomer Hipparchus in the second century BC. He called the brightest stars he could see first-magnitude stars (+1), those about half as bright he called second-magnitude (+2) and so on down to those barely visible which were sixth-magnitude.

When astronomers eventually reached the southern hemisphere they discovered brighter stars, so the scale had to be extended to magnitude 0 and then magnitude −1 and so on. Later, when the telescope was invented, the scale needed to be extended to +7 and beyond. The scale is referred to as the **apparent magnitude** (*m*) scale, shown in Figure 13.1.9 with some examples. On this scale, brighter stars have larger *negative* numbers and dimmer stars have higher *positive* magnitudes.

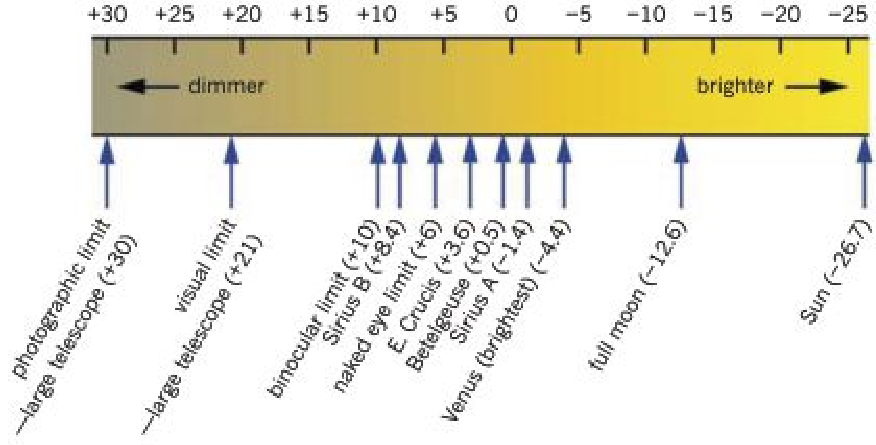


FIGURE 13.1.9 The apparent magnitude (*m*) scale is a combination of an ancient Greek scale with a nineteenth century mathematical redefinition. The visible stars range from −1.4 (Sirius A) 'down' to about +6 for stars barely visible to the naked eye in the best conditions.

In the nineteenth century astronomers were better able to measure the brightness of stars. They defined apparent magnitude (*m*) more precisely by saying that a difference in magnitude of 5 corresponds exactly to a factor of 100 times in apparent brightness (*b*). In other words it would take 100 magnitude +6 stars to equal the brightness of a single magnitude +1 star.

Mathematically, this means that each level of magnitude represents a change in brightness of about 2.5 times. In other words, it would take 2.5 stars of magnitude +6 to equal the brightness of a single +5 star. Or it would take 2.5 × 2.5 = 6.3 stars of magnitude +6 to equal the brightness of a +4 star, and so on.

PHYSICSFILE

Magnitudes

Magnitudes are effectively a logarithmic scale. A change of 5 magnitudes corresponds to a factor of 100. This is equivalent to the statement $x^5 = 100$. So $x = \sqrt[5]{100}$.
This means that an increase in the apparent brightness (*b*) of 1 on the magnitude scale corresponds to about $2\frac{1}{2}$ times the brightness. (Hipparchus' original scheme had this value at 2.)

Absolute brightness

Scientists knew there must be a relationship between the apparent brightness (*b*) of a star, its distance from Earth and its absolute brightness (*L*). Once the actual distances of a number of stars were known, it was possible to combine this with the apparent brightness to determine their absolute brightness (*L*). It was soon found that stars vary enormously in their absolute brightness (*L*).

For example, Sirius and Canopus, shown in Figure 13.1.10, are the two brightest stars in the sky. While Canopus appears almost as bright as Sirius, it is about 36 times further away than Sirius. To make up for this big difference in distance, it can be calculated that Canopus must be about 3000 times brighter than Sirius. On the other hand, the very closest star, Proxima Centauri, is not even visible to the naked eye and so must be a very dim star. Its absolute brightness (*L*) is only about 1/30 000th that of Sirius.

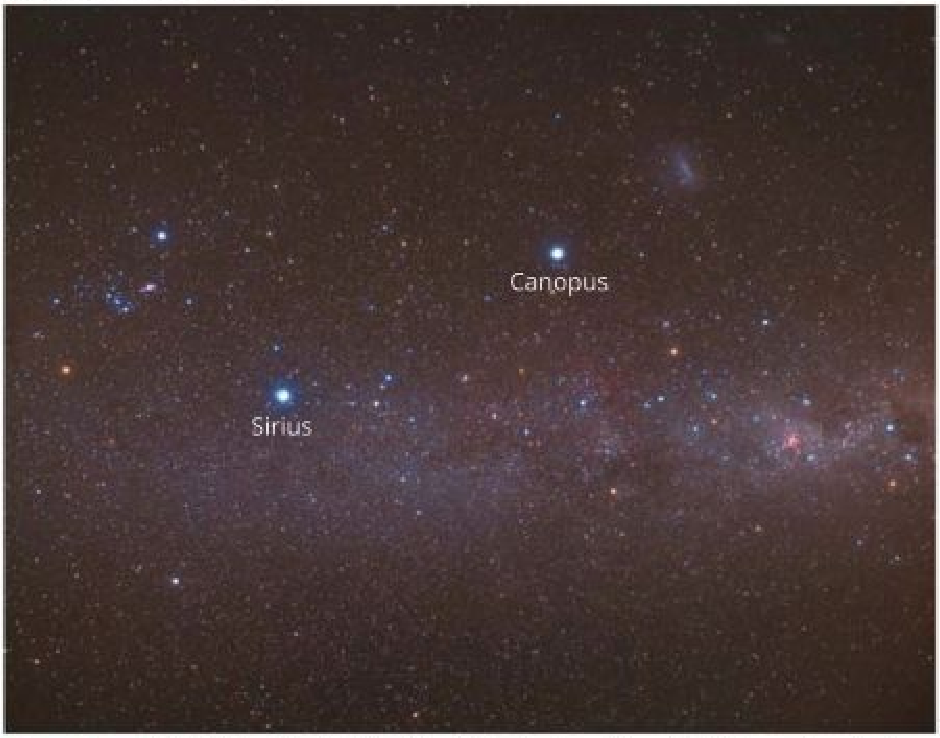


FIGURE 13.1.10 While Sirius is the brightest star in the sky viewed from Earth, Canopus is actually 3000 times brighter, but it is 36 times further away, so its apparent brightness isn't as great.

The absolute brightness (*L*) is often given on a scale of **absolute magnitude** (*M*), and is defined as being equal to the apparent magnitude (*m*) that the star would have if it was at a 'standard' distance of 10 parsec away from Earth. So the apparent magnitude (*m*) and absolute magnitude (*M*) of a star at 10 pc are the same. If the star is further away it will appear fainter and so the absolute magnitude (*M*) would be brighter than its apparent magnitude (*m*). If, for example, Canopus were at 10 pc instead of almost 100 pc, it would have an apparent magnitude (*m*) of –5.5. On the other hand, if Sirius were at 10 pc instead of 2.6 pc, it would be just an ordinary +1.5 star.

Finding distance from absolute and apparent magnitude

The absolute magnitude M of a star is related to its apparent magnitude m by the formula:

$$M = m + 5 - 5 \log d$$

where d is the distance of the star from the Earth, in parsecs.

This equation can be used to find the distance to a star if its absolute and apparent magnitudes are known. The apparent magnitude of a star can be measured from Earth. However, finding the absolute magnitude of a star can be more difficult. One way to do this is by making a comparison with a reference star that is similar and whose distance from Earth is already known. Another way is to use Cepheid variables. You may recall from Section 5.1 on page 168 that the brightness of a Cepheid variable changes slowly over a period of time. The absolute brightness of a Cepheid variable is related to the period of this variation. Based on this property, its brightness and also its absolute magnitude can be determined. Applying the equation means the distance to this star can be determined.

The following worked example illustrates how the equation can be used if M and m are known.

Worked example 13.1.4

FINDING DISTANCE FROM ABSOLUTE AND APPARENT MAGNITUDE

Sirius has an apparent magnitude of –1.5 and an absolute magnitude of 1.4. How far away is it?

Thinking	Working
Recall the rule connecting the apparent and absolute magnitude.	$M = m + 5 - 5 \log d$
Apply the rule to find the distance.	$1.4 = -1.5 + 5 - 5 \log d$ $\log d = 0.42$ $d = 10^{0.42}$ $= 2.6$ pc

Worked example: Try yourself 13.1.4

FINDING DISTANCE FROM ABSOLUTE AND APPARENT MAGNITUDE

A large distant galaxy has an apparent magnitude of 15 and an absolute magnitude of –20. How far away is it?

Brightness and luminosity

While the apparent magnitude (m) and absolute magnitude (M) scales are convenient for observational purposes, astrophysicists actually need to know the apparent brightness (b) and absolute brightness (L) in SI units. Apparent brightness (b) is measured in watts per square metre of received radiation, and absolute brightness (L) is measured in watts of total radiated power. When measured in this way the absolute brightness (L) is called the **luminosity** (L). When determining the luminosity of other stars, astronomers compare the apparent brightness (b) of the star with that of the Sun, then correct for the large difference in distance.

In this way astronomers were able to determine the luminosity of the thousands of stars for which stellar parallax has been measured. They have found an enormous range—from about 1/10 000 of the Sun's luminosity up to a million times. The Sun is a little below the middle in the range of luminosity, but the stars are not distributed evenly along this range. The great majority are actually dimmer than the Sun. Very luminous stars are the exception rather than the rule—only about 10% of stars are more luminous than the Sun.

13.1 Review

SUMMARY

- The speed of electromagnetic waves (light) in a vacuum is $c = 299792458 \text{ m s}^{-1} \approx 3.0 \times 10^8 \text{ m s}^{-1}$.
- While the speed of light is constant in a vacuum, when light travels from one medium to another its speed will change and the speed will differ depending on the density of the medium. The wavelength will also change. The frequency is a constant.
- The speed of light in a particular medium is $c = f\lambda$.
- Since frequency is the number of wavelengths per second, then the time in seconds for one complete wavelength to pass a given point is referred to as the period, T, and $T = \frac{1}{f}$.
- A spectroscope is a simple device combining a diffraction grating with a viewing telescope that allows the spectrum of light from a distant star to be viewed directly.
- A spectrometer is a spectroscope with a graduated scale enabling specific wavelengths revealed in the spectrum to be determined.
- The branch of science investigating the spectra produced when matter interacts with or emits electromagnetic radiation is referred to as spectroscopy.
- The distance to stars can be measured by the parallax movement they show as a result of the Earth's revolution around the Sun.
- The distance to over one hundred thousand stars has been measured by stellar parallax, but the vast majority of stars are much too distant for this method.
- The apparent brightness (b) of stars can be measured on a scale of apparent magnitude (m).
- The brightest stars are around –1 and those barely visible are around +6 on the apparent magnitude scale.
- The absolute magnitude (M) is equal to the apparent magnitude (m) that a star would appear to have if at a standard distance of 10 pc from us.
- The absolute brightness (L) is the luminosity of a star. It is the total power radiated. Other stars may have luminosities up to a million times greater or down to one ten thousandth that of the Sun. Most stars are less bright than the Sun.

KEY QUESTIONS

1 A particular wavelength of ultraviolet light has a wavelength of 350 nm. If the approximate speed of light is $3.0 \times 10^8 \text{ m s}^{-1}$, what is the frequency of this light correct to two significant figures?

The following information relates to questions 2 and 3.

Scorpius X–1 (Sco X–1) was the first X-ray source found in the constellation Scorpius. The X-ray emissions of Scorpius X–1 are 10000 times greater than its visual emission. It is the strongest source of X-rays in the sky other than our Sun.

2 If the wavelength of these emissions is an average of 8 nm, what is the frequency of the X-rays?

3 What is the corresponding period for the X-rays from Scorpius X–1?

4 What does the term redshift refer to?

5 A second is a very small fraction of a degree. How many seconds are there in 1°?

6 If a star shows a parallax of 0.44", how far away is it in pc and in m?

7 A star has a parallax angle of 1.56". Calculate its distance from Earth in parsecs.

8 61-Cygni has an apparent magnitude (m) of +5.21 and an absolute magnitude (M) of +7.49. What does this mean?

9 Betelgeuse has an apparent magnitude (m) of +0.5 and an absolute magnitude (M) of –7.2. What does this mean?

10 Rigel and Betelgeuse are actually about the same distance away, but their apparent magnitudes (m) are +0.12 and +0.50 respectively. Betelgeuse has an absolute magnitude (M) of –7.2. Which one of the following is the absolute magnitude (M) of Rigel?

A –6.5
B –7.0
C –8.1
D –9.5

11 A star has an absolute magnitude M of 4.83. At what distance will it have an apparent magnitude, m, of 14.83?

13.2 Classifying stars

Astrophysicists use spectroscopy to analyse light from distant stars. Spectroscopy involves spreading out the spectrum of light (both visible and invisible light) received from a star and then analysing patterns in the data collected. The spectra are compared with others from Earth that are known to scientists. From spectral analysis, the properties and even the life cycle of stars can be determined.

ANALYSING STAR LIGHT

A spectroscope is a fundamental tool for collecting spectral data from stars in the visible and near-visible range of the electromagnetic spectrum. Astronomers use a range of methods for collecting information that is not in the visible range to supplement the information they get from standard telescopes (see Figure 13.2.1).

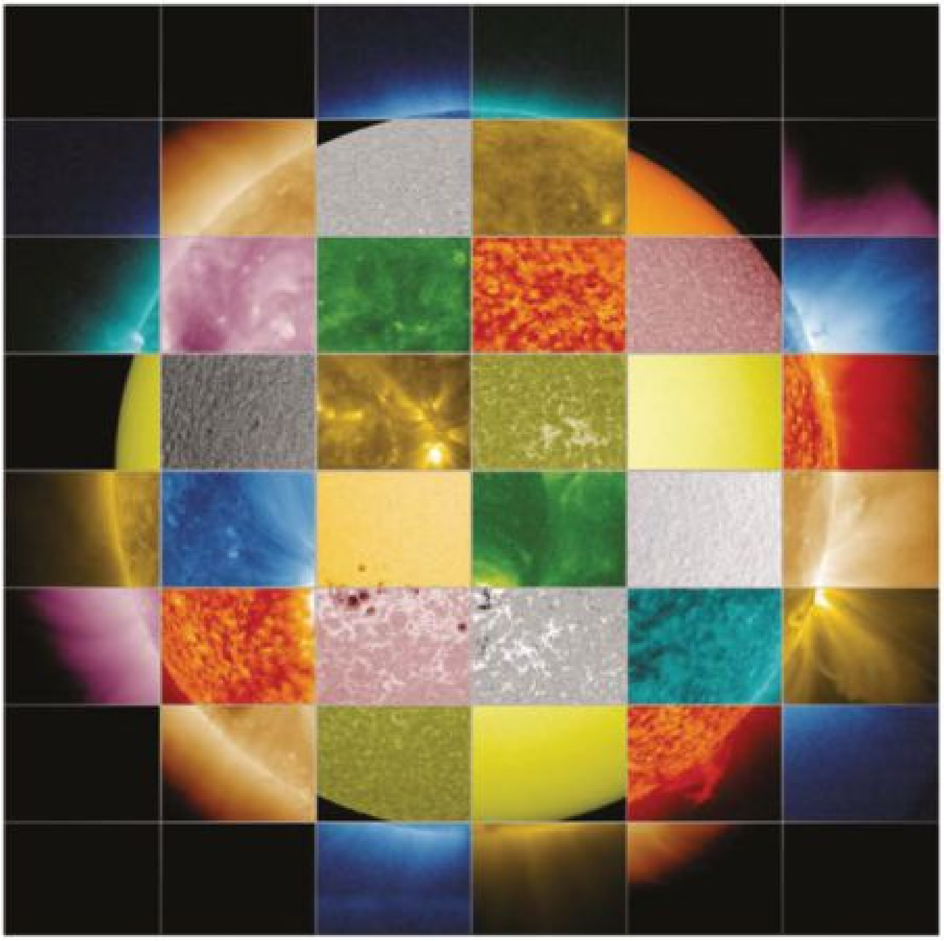


FIGURE 13.2.1 The Sun observed at different wavelengths within the electromagnetic spectrum. If you look at the Sun through a camera, it will look like a yellow circle because you are viewing visible light only. Viewing wavelengths outside the visible spectrum can reveal extra information about the Sun's surface and its atmosphere.

Radio telescopes, such as the one at Parkes in NSW, are used to collect the longer wavelength radio and microwave radiation. Because infrared radiation (IR) is absorbed by the Earth's atmosphere, satellites equipped with special IR telescopes which orbit above Earth's atmosphere are also used to photograph the sky at those wavelengths. While the atmosphere is transparent to wavelengths which are in the visible range or shorter than the visible wavelengths, the atmosphere still causes some distortion of these wavelengths. This distortion ultimately limits the power of Earth-based telescopes to detect these wavelengths accurately. For this reason, space-based telescopes such as the Hubble Space Telescope have proved to be of immense value to astronomers in detecting shorter ultraviolet and X-ray wavelengths.

You have seen in Chapter 2 'Applying thermodynamic principles' how hot objects emit radiation of different wavelengths. This emitted light is a continuous spectrum, but some wavelengths of light are more intense than others. The wavelength at which the peak intensity occurs gives an indication of the temperature of the surface of the object.

As stars are hot objects, they emit light across a range of wavelengths (that is, a spectrum). If the light from the star then passes through a cloud of gas, the gas will absorb certain wavelengths, leaving dark bands in the spectrum corresponding to the absorbed wavelengths. This process is shown in Figure 13.2.2. Figure 13.2.2(b) shows the dark bands. The absorbed light (in the cloud of gas) temporarily puts the atoms in the gas into an excited state. As the atoms return to the ground state, light is released. The re-emitted light radiates out in all directions, so if seen from the direction shown in Figure 13.2.2(c), it appears as bright bands. The bright bands form what is known as an emission spectrum. The bright bands of the emission spectrum will therefore correspond to the dark bands of the absorption spectrum.

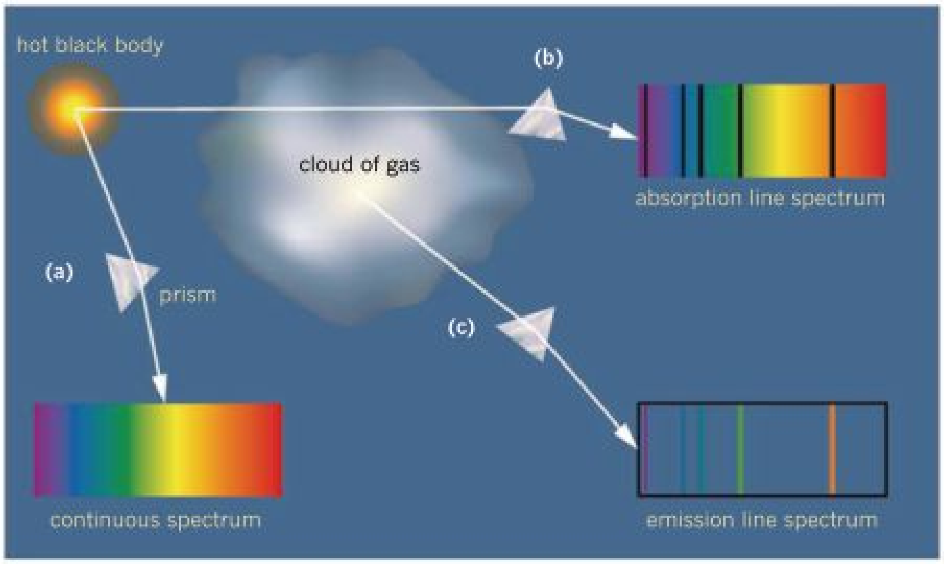


FIGURE 13.2.2 The hot object at the top emits a continuous range of wavelengths with the intensity dependent upon its temperature. (a) The spectrum of wavelengths can be seen when viewed through a prism. (b) If the light passes through a gas, some wavelengths are absorbed, giving rise to an absorption spectrum. (c) Re-emitted light from the gas produces an emission spectrum.

The lines emitted or absorbed are characteristic of each of the ninety-two chemical elements. For example, sodium always emits or absorbs two very characteristic lines in the yellow region. This explains why a barbeque flame turns yellow when you throw salt on the meat. If these two lines are seen in a spectrum it means that sodium is present. Whether it is present as part of a compound (salt is the compound sodium chloride) or as a vaporised gas makes no difference—the lines always have the same position in the spectrum. So, by examining the spectrum of a star, it is possible to tell what elements are present at its surface or in its environment.

The Sun's absorption spectrum has been examined in great detail and many elements have been identified. The dark lines in the solar spectrum are often referred to as Fraunhofer lines, as it was Joseph Fraunhofer who discovered them in 1817 (see Figure 13.2.3). In 1868, some lines were observed that did not seem to correspond to any element known on Earth. At the time this element appeared to be exclusive to the Sun and was called 'helium' (from the Greek word *helios*, meaning 'sun').

The absorption spectra from other stars are quite similar to the solar spectrum in most respects. This confirms that the basic chemistry and physics understood on Earth seems to apply throughout the universe. No strange new lines have ever been discovered in any other stellar spectra. There are subtle differences, however, and it is these differences that tell astronomers an enormous amount about other stars.

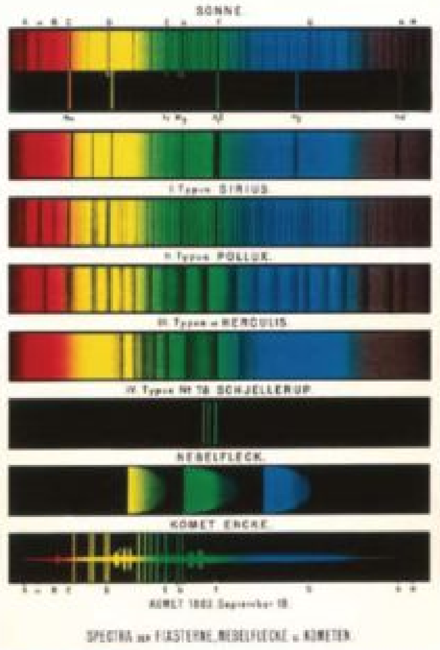


FIGURE 13.2.3 An historical 1884 diagram of the spectra of the Sun (shown as 'Sonne' in German) and other astronomical objects showing the absorption spectra known as Fraunhofer lines. The absorption lines are indicative of the presence of particular elements in the atmosphere of the stars.

You have already seen that the spread of colour through the continuous spectrum is an indication of the surface temperature of the star. The greater the proportion of blue and UV, the hotter the star.

In the 1890s, a scheme for classifying the spectra of stars was developed. It arranged the spectra in an order determined by the presence or absence of certain lines associated with hydrogen and labelled them alphabetically with the letters from A to O. It seemed to be a logical arrangement at the time, but later it became apparent that there was a better way to arrange them.

The new method involved placing similar spectra next to one another and finding a pattern which smoothly changed from one class of spectrum to the next over the whole range. As a result, some of the original classifications were dropped and the rest were re-ordered into a new sequence: OBAFGKM. Later modifications divided each class into smaller subdivisions by adding the digits 0–9 after the letters.

Table 13.2.1 below summarises the OBAFGKM classification.

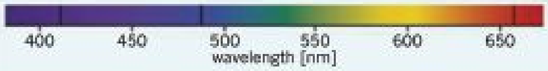
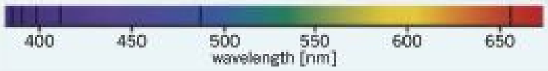
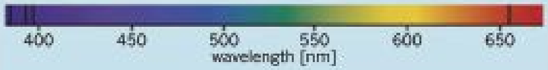
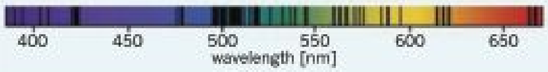
Spectral type	Approximate temperature (K)	Main characteristics	Spectrum	Colour
O	>30000	relatively few lines; the lines of ionised helium		blue
B	10000–30000	the lines of neutral helium		blue–white
A	7500–10000	very strong hydrogen lines		white
F	6000–7500	strong hydrogen lines; ionised calcium lines; numerous metal lines		yellow–white
G	5000–6000	strong ionised calcium lines; numerous strong lines of ionised and neutral iron and other metals		yellow
K	3500–5000	strong lines of neutral metals		orange
M	<3500	titanium oxide streaks		red

TABLE 13.2.1 The OBAFGKM classification of the spectra of stars.

The real significance of this pattern was not known at the time. It was later found that the changes in the spectra corresponded to differing temperatures, with the O stars being the hottest and the M stars being comparatively cooler. It was shown that the changes between the classes were associated with the fact that different atoms become ionised (lose electrons) at various temperatures. At cooler temperatures the light may not have enough energy to excite the atoms sufficiently to create some of the lines. So it was realised that lines may appear over a certain temperature range but not at higher or lower temperatures.

This discovery provided a new way to determine the temperature of a star which was not affected by colour change as light travelled through interstellar gas. The difference between the two methods gave astronomers useful information about the interstellar gases that might be present between the star and the Earth.

A careful analysis of a spectrum reveals much important information about the star:

- its temperature
- the elements that are present and the state in which those elements exist (solid, liquid, gas)
- the pressure of the gas emitting the light
- any magnetic fields present.

The radius of a star can also be determined with the help of the spectrum. The diagram in Figure 13.2.4 sums up the relationship between some of these measurements and quantities.

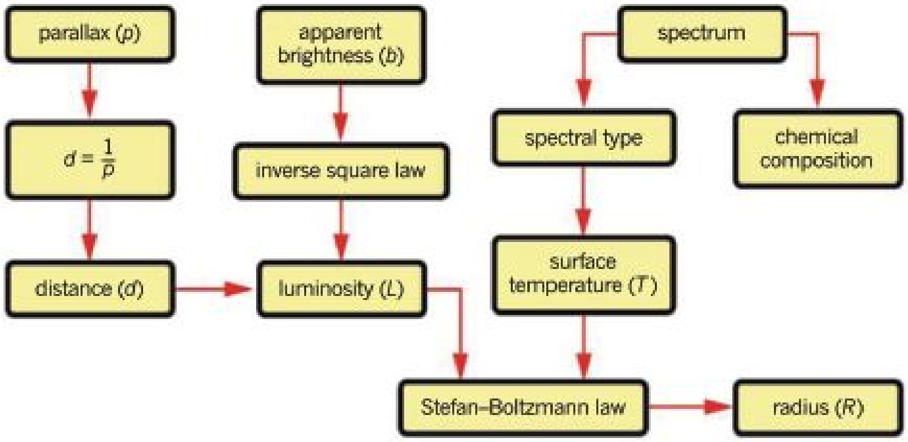


FIGURE 13.2.4 A mind map illustrating what particular aspects of star light can tell astronomers about a star.

If the surface temperature of a star is known from the spectrum, a basic law of physics (the Stefan–Boltzman law discussed in Section 2.1) can be used to determine the amount of energy given off each second by each unit area of the surface. Knowing this and the luminosity (the total energy given off by the star) enables the total surface area and hence the radius of the star to be calculated. The luminosity is found by comparing the apparent brightness with the distance determined by stellar parallax.

THE HERTZSPRUNG–RUSSELL DIAGRAM

Once data for many stars had been collected it was natural to look for patterns in the data. In 1911 and 1913, the astronomers Ejnar Hertzsprung in Denmark and Henry Russell in America independently discovered a pattern which brought order into the apparent chaos. The graph they produced is now known as the **Hertzsprung–Russell diagram**, or H–R diagram for short. In effect they plotted the luminosity of a star against its temperature and found a distinct set of patterns, as shown in Figure 13.2.5 on page 469.

Figure 13.2.5 shows the common way of arranging the H–R diagram with luminosity on the vertical axis (brighter at the top) and temperature on the horizontal axis. Each axis is labelled with two different scales. Temperature is derived from spectral type, so these are both plotted on the horizontal axes. Luminosity is derived from the absolute magnitude, so these are plotted together on the vertical axes.

You'll notice from the H–R diagram that the stars are not randomly distributed. The band of stars from the lower right corner to the upper left corner is called the **main sequence** and includes 90% of the stars in the sky. Main sequence stars go from dull, cool stars at the bottom right to bright, hot stars at the upper left of the diagram. There is a clear relationship between the brightness and temperature of a star. Generally, it was found that hot stars in the top-left corner of the H–R diagram are brighter and bigger, but there were important exceptions to this rule as you can see.

In one sense this is hardly surprising—if two stars are the same size it is reasonable to expect the hotter star to be brighter. However, stars in the upper right of the diagram, such as Betelgeuse, are very large. They need to be large in order to be so bright, given that their temperatures are relatively low. This group of stars are called **giants** or supergiants and are over one thousand times the size of the Sun (see Figure 13.2.6 on page 469). On the other hand, stars in the opposite lower left corner must be small as they are not bright, despite being very hot. These stars are only about one thousandth the size of the Sun and are called **white dwarfs**.

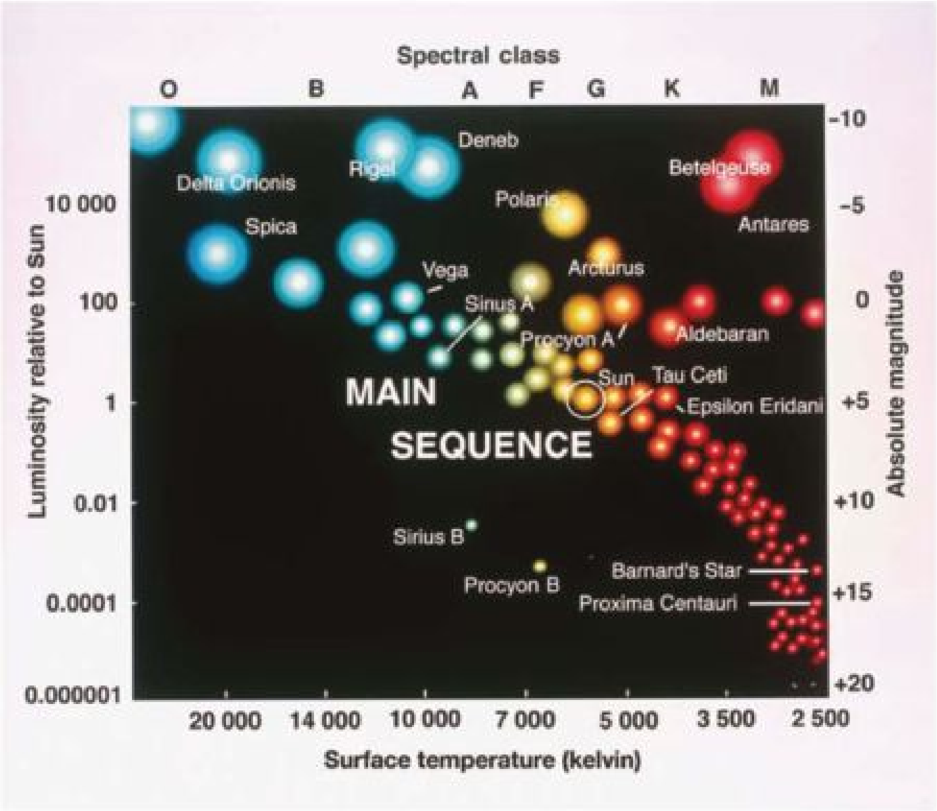


FIGURE 13.2.5 The Hertzsprung–Russell (H–R) diagram is one of the astronomer's basic tools. The luminosity scale takes the Sun's value as 1, which corresponds to an absolute magnitude of about +4.8. The temperatures corresponding to the spectral types are plotted along the bottom. Hotter stars are to the left, indicated also by the colour.

Most stars are on the main sequence, and the hotter they are the brighter they are. The Sun is approximately in the middle of the main sequence. You can see from the H–R diagram that most stars are both cooler and dimmer than the Sun—there is a greater density of stars on the lower right of the main sequence. Only about 1% of stars are giants or supergiants, and a further 9% are white dwarfs.

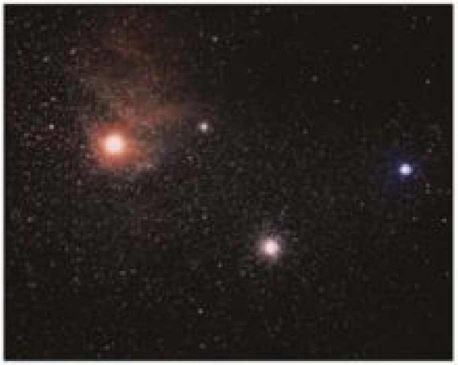


FIGURE 13.2.6 The brightest star, at centre left, is Antares (alpha Scorpii), a red supergiant star 500 light-years from the Earth.

Placing stars on the H–R diagram

If a new star is found, it is placed on the H–R diagram. Provided it is within stellar parallax range (remember that the satellite Hipparcos measures out to less than 1000 pc), luminosity and temperature can be measured directly. If the new star is beyond parallax range, its temperature can still be determined from its spectrum but its luminosity can't be determined directly. The spectral lines provide additional information that allows luminosity to be determined.

Spectral lines in stellar spectra tell us more than just the elements present.

- Magnetic fields tend to 'split' lines slightly.
- The pressure of the gas from which the light originated also has a subtle effect on some lines but not on others.
- Certain lines in the hydrogen spectrum are sensitive to the number of collisions the hydrogen atoms are experiencing. The collisions tend to slightly increase or decrease the energy and hence the wavelength of the emitted light. This in turn slightly broadens the spectral line. A giant star will tend to have a less dense photosphere, and hence the atoms will experience fewer collisions. This means that the spectral lines are less broadened than those of a main sequence or dwarf star.

Analysing the more subtle characteristics of the spectrum enables astronomers to classify the star and hence place it in the appropriate category on the H–R diagram.

Once the new star is placed on the diagram its luminosity can be determined. Given its luminosity and apparent brightness, the distance to the star can be calculated. This method enables astronomers to determine the distances to most of the stars in the sky, instead of just the small number within stellar parallax range. The technique of determining distance from the spectrum, rather than from parallax, is given the rather misleading name **spectroscopic parallax** (see Figure 13.2.7). The word parallax is used because it refers to a process of determining distance, but it is not to be confused with the process which uses genuine parallax called stellar parallax.

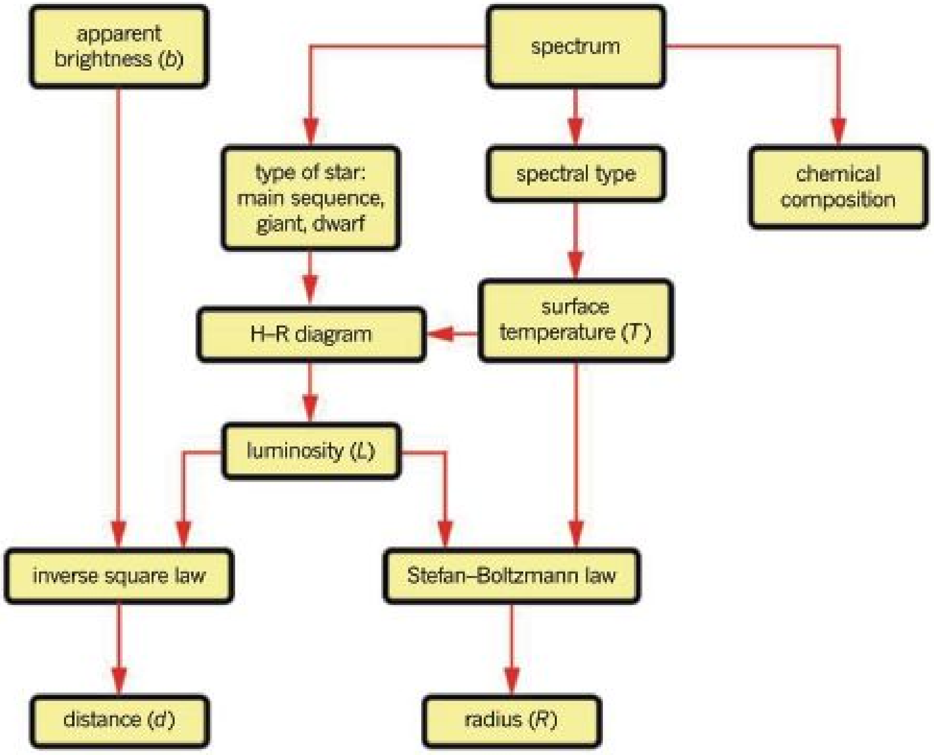


FIGURE 13.2.7 Spectroscopic 'parallax'. Knowing the apparent brightness and correctly interpreting the stellar spectrum to decide whether a star is a main-sequence, giant or dwarf star enables astrophysicists to deduce most of the important characteristics of a star.

Worked example 13.2.1

DETERMINING THE BRIGHTNESS OF A STAR

It is found that two stars have the same radius and are the same distance from Earth. One looks red and the other blue. Which one will be brighter?

Thinking	Working
The colour is an indication of the surface temperature of the star, hotter stars being bluer.	The blue star must be hotter.
Because it is hotter, the blue star will be radiating considerably more energy.	The blue star will appear significantly brighter.

Worked example: Try yourself 13.2.1

DETERMINING THE BRIGHTNESS OF A STAR

It is found that two stars have the same radius and are the same distance from Earth. One looks blue and the other yellow. Which one will be brighter?

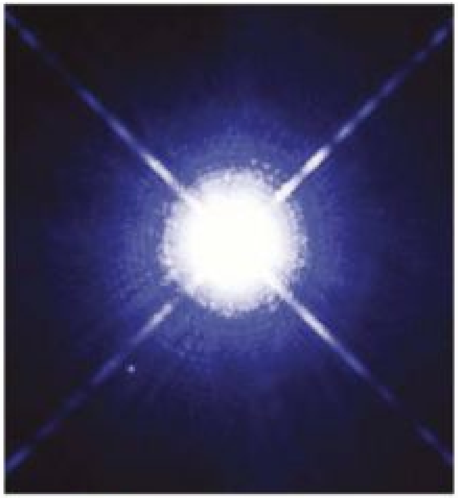


FIGURE 13.2.8 A Hubble telescope image of Sirius, the brightest star in the night sky and its companion, Sirius B. Sirius B is the small, round object to the bottom left of Sirius and was the first white dwarf to be discovered. The changes in their relative positions enables the orbits of the stars around each other to be determined and hence their respective masses can be found.

DETERMINING THE MASS OF A STAR

A lot can be determined about a star from its apparent brightness and spectrum. But one important quantity does not appear on that chart: the star's mass.

The Sun's mass was effectively determined by studying its gravitational effect on the planets. If the Sun suddenly lost half its actual mass, the Earth would not be held as strongly to the Sun and so would move off into a bigger and slower orbit, or would even escape from the Sun's gravitational field altogether.

In order to tell the mass of any star, something must be observed in orbit around it—the faster and closer the orbit, the heavier the star must be. Most stars do have planets, so this is one means that can be used where the star is not too distant for the gravitational effects on the orbits of the planets to be observed.

An alternative method involves only the stars themselves. Many stars are actually binary stars, that is, two stars revolving around their common centre of mass. Sirius and Sirius B, shown in Figure 13.2.8, are examples of binary stars. If the two stars can be seen as separate objects (and many of the closer binary stars can) their orbits can be found. These orbits will depend on the masses of the two stars. Heavy, close stars would revolve around each other quickly while less-massive stars or those separated by greater distances would move more slowly.

If the distance to a star is known, the size of the orbit can be calculated. Finding the period and size of the orbits is a matter of observation over many years. If the stars are within parallax range, Newton's laws of motion and of universal gravitation then enable the determination of the mass of both of the stars involved.

Outside of stellar parallax range, small shifts in the spectral lines of some stars show that they are revolving around a companion star or stars. Careful analysis of these so-called spectroscopic binaries enables many more determinations of mass.

The mass of a star is also closely related to its luminosity. Along the main sequence, stars with greater luminosity have greater mass. This can be seen in Figure 13.2.10. In this diagram, the symbol $M_{\odot}$ refers to the mass of the Sun. By comparing this diagram with the H–R diagram, you will discover that many stars are lighter than the Sun.

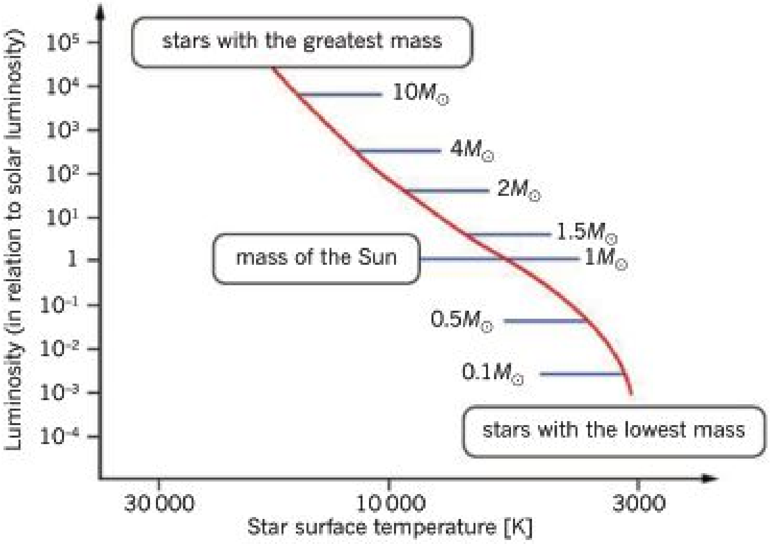


FIGURE 13.2.10 The majority of stars shine more weakly and are lighter than the Sun. There are a few huge stars, however, with masses which are more than ten times greater than the Sun.

It turns out that the luminosity of a star varies roughly with the mass cubed (that is, $L \propto m^3$). So a star twice the mass of the Sun would be about eight times brighter. There are exceptions to the rule, however. Giants and white dwarfs do not obey this relationship. The fact that all main sequence stars seem to conform to the same pattern confirms that they are basically similar in composition and are powered by the same reactions in their core. We will look at this in more detail in the next section.

PHYSICSFILE

Spectroscopic binaries

Spectroscopic binaries are pairs of stars rotating around each other, which are too far away to be distinguished separately in a telescope. Their spectra provides the information through which they can be identified as binaries.

In some cases what appears to be one star has spectral lines that usually appear in two different types of stars but never in the same one. More particularly, if two stars are rotating, it's likely that one is moving towards the Earth and the other away. In this case the lines from the star moving towards Earth will be blueshifted and the one moving away will be redshifted. This enables a determination of the relative speeds of the stars. Over time, changes in the spectra enable the periods of the orbits to be determined.

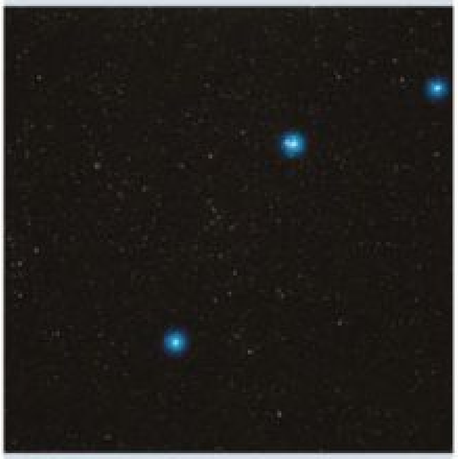


FIGURE 13.2.9 These three stars form the handle of the group known as The Plough, or Big Dipper, part of the constellation Ursa Major, the Great Bear. At centre right is Mizar (Zeta Ursae Majoris). Mizar is a spectroscopic binary star. Its companion, Alcor, is seen just to its left.

13.2 Review

SUMMARY

- The continuous spectrum of a star provides information about the temperature of its surface and the absorption lines tell us what elements are present and under what conditions.
- The spectra of stars are quite similar to the spectrum of the Sun, indicating that they have similar characteristics and processes.
- Stellar spectra are classified according to a system which grades them by surface temperature (OBAFGKM; hot to cool respectively).
- The Hertzsprung–Russell (H–R) diagram plots stars according to temperature and luminosity.
- All stars fall into three main groups: main sequence; giants and supergiants; and white dwarfs.
- Moving up the main sequence on the H–R diagram, the stars become hotter, brighter and larger.
- The mass of binary stars can be determined from the period of motion and the size of their orbit.
- For main sequence stars, the luminosity varies roughly with the mass cubed.

KEY QUESTIONS

1 Which properties of a star are plotted on a Hertzsprung–Russell diagram?

2 According to the H–R diagram, how does the luminosity of a star relate to its temperature?

3 What information do the continuous spectrum emitted by a star and its absorption spectrum reveal about the star?

4 How does the H–R diagram help astronomers to determine the distance to a star which is too far away for stellar parallax?

5 A new star has been discovered and astronomers wish to place it on the H–R diagram. It's observed that the spectral lines on the spectra of the star are significantly less broadened than those from the Sun and the spectrum has a peak intensity towards the red end of the visible spectrum. What type of star has been observed?

6 Referring to the H–R diagram, order the following stars from largest to smallest:

Betelguese
the Sun
Sirius A
Proxima Centauri
Sirius B
Vega

7 Why is the determination of a star's surface temperature crucial to finding the size of a star?

8 Main sequence stars are known to obey a mass–luminosity relationship. If a star is found to have a luminosity eight times that of the Sun, what is its mass likely to be (in terms of the mass of the Sun, $M_{\odot}$)?

A $8\ M_{\odot}$
B $4\ M_{\odot}$
C $2\ M_{\odot}$
D $1\ M_{\odot}$

9 Which of the following best describes what is plotted on the Hertzsprung–Russell diagram?

A absolute magnitude against the luminosity of the stars
B apparent magnitude against the temperature of the stars
C temperature against spectral type of the stars
D luminosity against spectral type of the stars

10 In general, which of the following is true for most stars?

A luminosity increases with the surface temperature
B luminosity decreases with the surface temperature
C luminosity is not related to the surface temperature
D luminosity is directly proportional to the surface temperature

13.3 The life and death of stars

Before the twentieth century it had generally been thought that the stars were basically permanent features of an unchanging eternal universe. Scientists now understand the processes that fuel the nuclear reactions inside stars. As this understanding has developed, it has become obvious that while stars last a very long time, they do not last forever. The Sun, for example, will not last forever. The changes it will experience through its life cycle are depicted in the artwork in Figure 13.3.1.

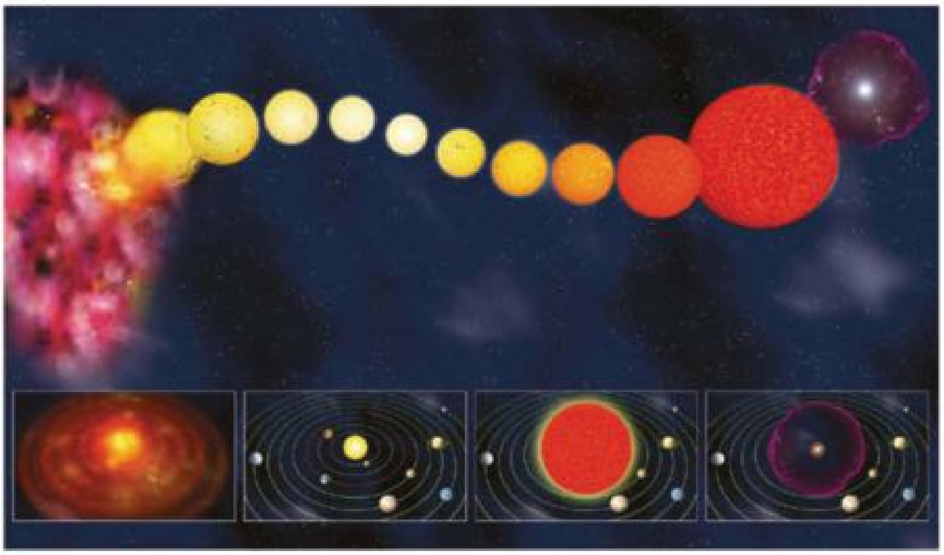


FIGURE 13.3.1 The life cycle of a star, such as the Sun, from molecular gas cloud in the beginning to the predicted red giant which will eventually swallow up the inner planets, to a white dwarf.

The story of how astrophysicists have come to interpret the H–R diagram and to build up a picture of the life and death of stars is one of the most fascinating in physics.

THE SUN: EARTH'S CLOSEST STAR

It was only after the time of Galileo and Newton that scientists realised that the stars were actually Sun-like objects a very long way away. The best way to learn about stars, then, was to look at the closest one.

Galileo's discovery of sunspots led to the understanding that the Sun rotates on its axis (see Figure 13.3.2). However, the equator of the Sun was seen to have a period of about 25 days while regions of higher latitude took several more days for a full cycle, indicating that the Sun is not a solid body like the Earth. Because it is obviously so hot it was assumed to be gaseous, but just how it produced so much heat wasn't understood.

It was known that life had existed on the Earth for at least several hundred million years and so that meant that the Sun must have been radiating energy at fairly much the same rate for at least that time. The problem was that any known mechanism for producing heat could not possibly have generated so much energy over such a long period of time.

The English physicist Lord Kelvin (who also determined the temperature of absolute zero) and the German physicist Hermann von Helmholtz (Figure 13.3.3 on page 474) proposed that vast amounts of heat would be generated from the enormous weight of gas collapsing into the Sun. As it fell, the potential energy would be converted into kinetic energy and would produce heat. This was good physics, but again it could not possibly last for the billions of years that the Sun has been producing energy. Scientists now know their theory was wrong, but the Kelvin–Helmholtz contraction does turn out to be important in the formation of new stars—and in the death of old ones.

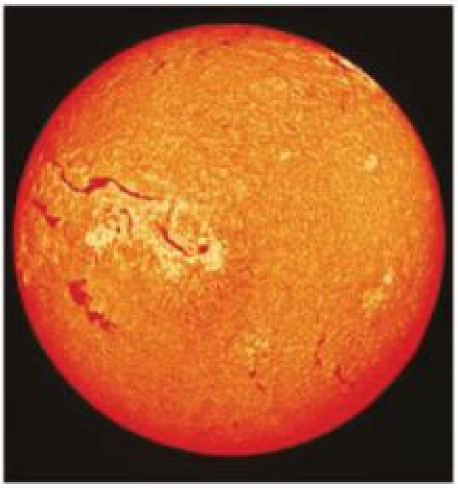


FIGURE 13.3.2 The apparent movement of sunspots and other features across the Sun's surface make its rotation clear. Sunspots and flares reveal the active and on-going processes within the Sun.

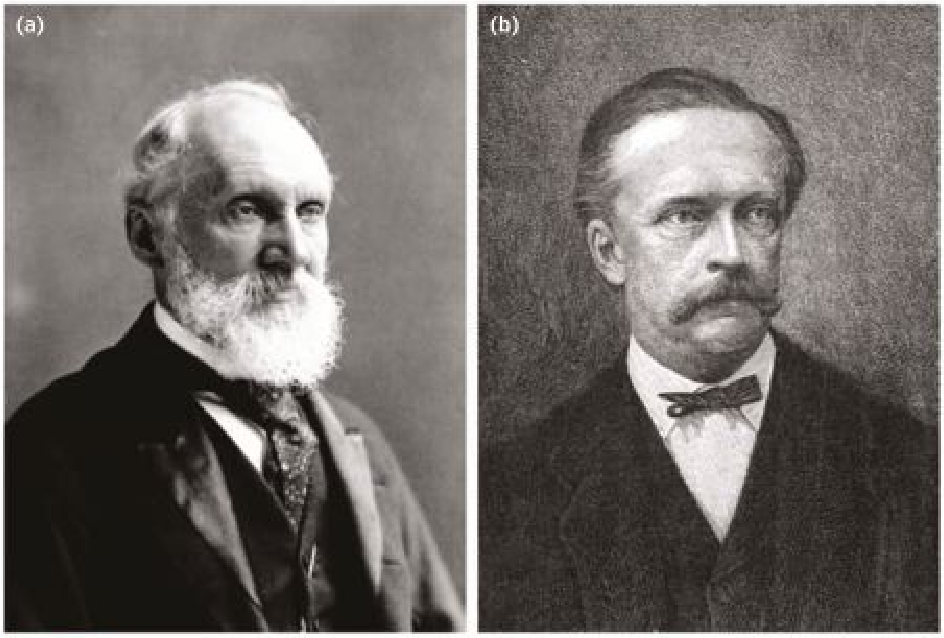


FIGURE 13.3.3 (a) Lord Kelvin, 1824–1907; (b) Hermann von Helmholtz, 1821–94.

Calculations based on the mass and energy output of the Sun showed that the amount of energy being produced for each atom in the Sun was around a hundred million times greater than the energy produced by each atom in chemical reactions. Something very different was going on in the Sun.

The clue came with Einstein's theory of special relativity. Einstein's equation $E = mc^2$ is well known, but the real meaning of this equation is not that simple. A reasonable interpretation for our present purposes is that there is a huge amount of energy locked up in mass. The factor c^2 is the speed of light squared and so is a very large number (approximately 9×10^{16} m^2 s^{-2}). This means that if this energy can be released, a small loss of mass could produce a huge amount of energy.

It was thought that the Sun was somehow tapping into this so called 'mass–energy'. There's an enormous amount of energy being radiated from the Sun. Assuming the energy was being produced mainly in the core, it was calculated that the temperature at the centre of the Sun must be millions of degrees. This would mean that the atoms at the centre would be stripped of their electrons and there would be just a frenzied mass of nuclei and electrons flying around at enormous speeds. Since the density of the Sun was relatively low, it was assumed that it must be composed mostly of hydrogen with perhaps some helium.

In the 1920s, British astrophysicist Robert Atkinson put all of this together and suggested that in the conditions at the centre of the Sun the hydrogen nuclei may 'fuse' together, creating helium nuclei. If this were the case, huge amounts of energy would be released—this is nuclear fusion as described in Chapter 7 'Energy from the atom'.

PHYSICSFILE

Einstein's $E = mc^2$

It is often stated that in nuclear reactions 'mass is converted into energy'. This is actually an oversimplification of Einstein's ideas. There is no mass 'shaved off' the particles and somehow mysteriously turned into energy. What is true is that the total mass of the particles in a helium nucleus is a little less than the total mass of the equivalent particles in hydrogen nuclei. They are exactly the same particles as before, except that they now have a little less total potential energy than before.

Einstein showed that mass is actually a property not of the individual particles, but of the system of particles including the energy bound up in the forces between them. As some of the energy has been released, so also has the equivalent mass.

Worked example 13.3.1

MASS–ENERGY

The Sun is producing about 4.0×10^{26} J of energy every second as visible and invisible radiation. At what rate is the Sun losing mass due to this total energy loss? (Use $c = 3.0 \times 10^8$ m s^{-1}.)

Thinking	Working
The energy comes from the fusion of hydrogen into helium with a corresponding loss in the potential energy of the nuclei. This loss of energy will correspond to a mass loss given by Einstein's equation $E = mc^2$.	$E = mc^2$ $E = 4.0 \times 10^{26}$ J $c = 3.0 \times 10^8$ m s^{-1} $m = ?$
Rearrange $E = mc^2$ in terms of mass and solve.	$E = mc^2$ $m = \frac{E}{c^2}$ $= \frac{4.0 \times 10^{26}}{(3.0 \times 10^8)^2}$ $= 4.4 \times 10^9$ kg s^{-1} So the Sun is losing mass due to visible and invisible radiation at a rate of 4.4×10^9 kg s^{-1}.

Worked example: Try yourself 13.3.1

MASS–ENERGY

The visible portion of the energy the Sun is producing each second is approximately equal to 5.0×10^{25} J s^{-1}. At what rate is the Sun losing mass due solely to this energy loss?

(Use $c = 3.0 \times 10^8$ m s^{-1}.)

Fusion in the Sun

The combination of Einstein's theory of relativity and an improved model of the atom (Bohr's quantum atom was suggested in 1913) enabled astrophysicists to understand the processes occurring in the Sun. Today, computer models have enabled considerable insight into the mechanisms that keep the Sun 'burning' in the sky.

The basic principle used in modelling the Sun is that any part of it must be in what is called **hydrostatic equilibrium**. That means that on any 'piece' of Sun, the inwards pressure from the weight of all the material above it must be balanced by the outwards pressure of the radiation released by the nuclear reactions in the core below.

These models suggest a number of theories about the Sun's structure and production of energy. Nuclear fusion occurs in a zone that extends from the centre out to about 0.25 of the Sun's radius ($R_{\odot}$). The temperature in this region is above 10 million degrees, the density is about 160 000 kg m^{-3} and the pressure is about 340 billion times the Earth's atmospheric pressure.

The energy flows outwards from this zone by a combination of convection and radiative diffusion. Radiative diffusion is a process in which light bounces around, transferring heat in the process. It is the main mechanism for energy transfer out to about $0.7R_{\odot}$, where convection takes over. At this point the temperature is down to around one million degrees and the protons and electrons come together to form hydrogen atoms, which absorb the light more effectively and so the radiative diffusion becomes less effective. The density is only 80 kg m^{-3} (much less than water, but about 60 times that of air) and the pressure is down to about 10 million 'atmospheres'. Most (99%) of the Sun's mass is below this level.

At the surface, the temperature is down to approximately 5800 K. There is actually no real 'surface', just a layer where the churning hot gases start to sink again as they lose their energy by radiation out into space. This layer, called the **photosphere**, is a thin layer from where the Sun's visible light is emitted. The whole process may sound rapid but it has been calculated that energy produced by the nuclear fusion in the Sun's core takes about 170 000 years to travel out to the surface. It then takes just over eight minutes to reach the Earth. The energy received from the Sun today was actually generated inside the Sun when humans were just beginning to diverge from the apes!

Some properties of the Sun are listed in Table 13.3.1.

Average distance between the Earth and the Sun (1 AU)	1.496×10^{11} m (390 times as far as the Moon)
Angular diameter from Earth	0.5° (same as the Moon)
Mass of the Sun	1.99×10^{30} kg (300 000 times the Earth)
Diameter of the Sun	1.4×10^{9} m (109 Earth diameters)
Surface temperature	5780 K
Average density of the Sun	1.4×10^{3} kg m^{-3} (Earth is 5.5×10^{3} kg m^{-3})
Energy output of the Sun	3.86×10^{26} W (joules per second)

TABLE 13.3.1 The physical properties of the Sun.

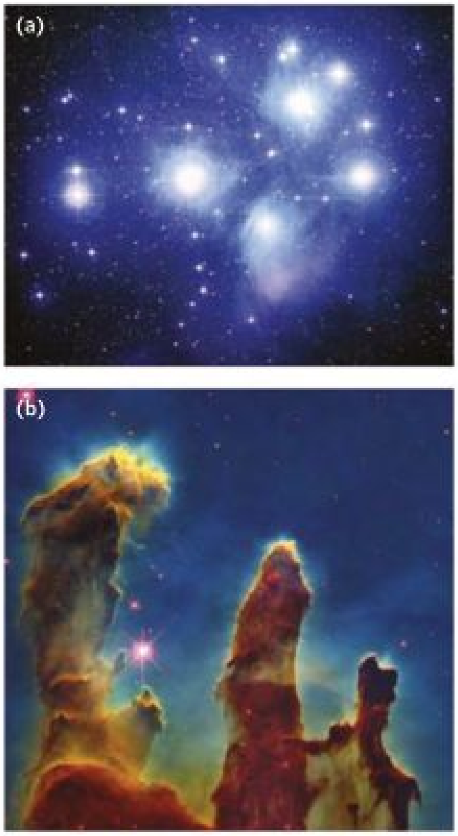


FIGURE 13.3.4 (a) The Pleiades star cluster, also known as the 'seven sisters' or 'M45'. It includes over 500 young stars, including the seven brightest visible to the naked eye. It can be seen in the constellation 'Taurus'. (b) This famous photograph of the Eagle Nebula taken by the Hubble telescope shows huge gaseous pillars several light-years high. The pillars are evaporating gaseous globules and are thought to be regions in which stars are forming. The nebula can be found near the constellation of Sagittarius.

STELLAR EVOLUTION

At the rate at which the Sun is consuming its hydrogen, it should run out of fuel in about a hundred billion years. However, models of the nuclear processes predict a much shorter life.

As the heavier helium nuclei build up in the core of a star, the nuclear fusion reaction zone moves outwards. Once around 10% of the hydrogen is consumed, the star becomes unstable and the outer layers expand until the star becomes about ten times the size. In other words the star becomes a giant. The Sun is due to do this in around six billion years and, when it does, it will engulf the Earth.

For this reason it was originally thought that the stars are 'born' on the main sequence of the H–R diagram and then eventually moved into the giant phase. Heavier stars were expected to burn their fuel more rapidly than lighter stars. Although they have more fuel, computer models predict heavier stars will run out of fuel sooner and so they should have a shorter lifespan.

Confirming this model can't be achieved by watching what happens to a particular star through its lifetime as the time scale is potentially huge. Scientists have to infer, from the different types of stars that can be seen, at what stages of their lifetime they are. Then they need to check whether this information fits the theories.

There are groups of stars, called star clusters, which seem to have all been born at about the same time out of the same cloud of 'dust' and gas. The Pleiades group, or seven sisters as it is more commonly known, is one such cluster. The famous 'Starbirth' photograph shows another cloud where clusters of stars could be forming. Images of both are shown in Figure 13.3.4.

These clusters provide the opportunity to look for differences between stars of similar ages but different sizes. If the modelling that suggests larger stars will have shorter lifetimes is correct, there should be evidence of this in clusters. The bluer, hotter and larger stars should be moving off the main sequence and becoming giants sooner than the smaller, cooler stars.

Many clusters were studied in this way in the 1950s and this is just what was found. Stars at the upper end of the main sequence were more likely to have moved towards the giant area than those further down.

The number of stars seen in the various areas on the H–R diagram is presumably an indication of the amount of time stars spend in that area. Because there are many more stars in the main sequence than in the giant phase it is assumed that stars spend most of their life as reasonably stable main-sequence stars. There are few stars between the main sequence and the giant phase and so it can be deduced that after reaching the end of their main sequence life, they fairly rapidly expand to become red giants.

The life cycle of a star

With the aid of computer modelling, a fairly complete picture of the birth, life and death of stars has been built up that agrees well with observational evidence.

A star begins its life as a **protostar**, a large mass of gas and dust that has come together as a result of gravitational attraction. The heat generated in this gravitational collapse (the so-called Kelvin–Helmholtz contraction mentioned earlier) causes the protostar to become very hot. As the gravitational collapse continues, the gravity becomes stronger, and accelerates the collapse even further (Figure 13.3.5).

FIGURE 13.3.5 The Horsehead nebula in the constellation Orion. This nebula is part of the Orion nebula complex, an enormous starbirth region some 1500 light-years from Earth.

Eventually the interior of the new star becomes hot enough, at about 10 million kelvin, to ignite the nuclear fusion reactions. The extra heat generated by these reactions eventually stops the gravitational collapse. This is as a result of the outwards pressure of the radiation from the hot gas. At this point the protostar has become a new main-sequence star and remains in a fairly stable condition for a time that is dependent on the size of the star. This time will be measured in millions of years for very massive stars, billions of years for stars like the Sun, and tens of billions of years for the majority of stars lower down the main sequence.

Once a star becomes a red giant a new process starts to occur in the core. The further gravitational collapse of the heavy core heats it to even higher temperatures. At around one hundred million degrees, new nuclear reactions start to occur: three helium nuclei combine to form carbon. This reaction more or less stabilises the giant star in its position on the H–R diagram.

Eventually, however, the nuclear fuel begins to run out and the star begins to contract, a process that can keep it hot and actually raise its temperature so that it moves back towards the left of the H–R diagram. The overall brightness decreases and it gradually moves down the left side of the diagram to the region of white dwarfs. This journey through the H–R diagram is plotted in Figure 13.3.6.

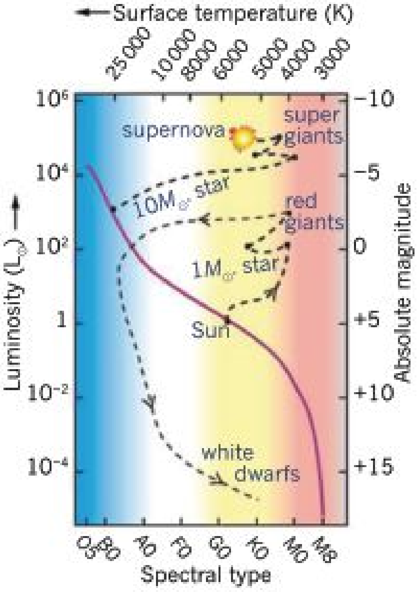


FIGURE 13.3.6 Once stars reach the end of their life on the main sequence they move up into the giants. Depending on their size, they may fade to a white dwarf or become a supernova.

At this stage, the star has collapsed into electron degenerate matter that basically means that the atoms have collapsed into a mass of protons, neutrons and electrons with a density about a million times that of water. They gradually radiate their remaining heat away, become cooler and redder, and slide down and out of the H–R diagram at the lower right as **black dwarfs**.

This picture of stellar evolution is appropriate for most stars, but a more spectacular fate awaits some.

A **supernova** is the explosive end of a massive star. There have been a number recorded in history. There was a notable supernova in the constellation Cassiopeia in 1572 that was observed by Tycho Brahe, who was so fascinated he went on to become one of the great astronomers of the period. A 1604 supernova had a similar effect on another great astronomer, Johannes Kepler.

In 1987, a new supernova (shown in Figure 13.3.7) suddenly appeared in the Large Magellanic Cloud, a fuzzy area near the Southern Cross. The Large Magellanic Cloud is actually a satellite galaxy of the Milky Way.

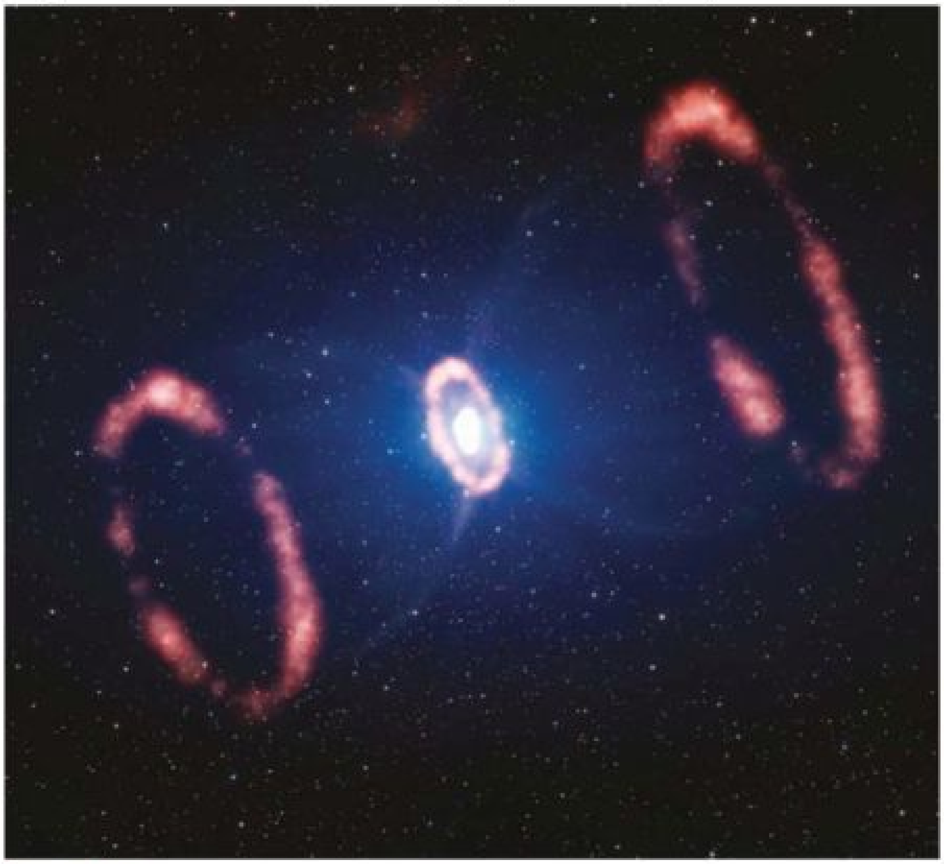


FIGURE 13.3.7 The remnants of Supernova 1987a. The rings are thought to be glowing gas that was ejected when the star was a red giant but which was ionised by the intense UV light from the explosion.

Huge pressure is created inside a star with a mass of around four times the Sun's mass ($4M_{\odot}$). After the initial hydrogen-burning phase, the temperatures created by the contracting star can reach 600 million degrees, at which point new nuclear fusion reactions can take place. In these reactions, the carbon produced by the helium fusion is 'burnt' to produce oxygen, neon, sodium and magnesium.

In even more massive stars, various heavier elements are produced and burnt in a series of reactions that occur at temperatures up to billions of degrees. The effect of these reactions is to move the star around on the H–R diagram as it changes brightness and temperature with each new set of reactions. Stars over about $25M_{\odot}$ eventually become so large and bright that they form the supergiants at the top right of the H–R diagram. Betelgeuse and Rigel in Orion are examples of these supergiants.

The age of a star can be determined by this change in chemical composition. Young stars will consist largely of the hydrogen from the gas clouds from which they formed and the helium into which the hydrogen is fused. Their emission spectra will show this chemical characteristic. Older stars will exhibit spectra with successively higher levels of the heavier elements created as the fusion process continues.

Clusters of stars and even galaxies of similar-aged stars will tend to produce light that exhibits the emission spectra characteristics of the elements being produced in the individual stars. The absorption by hydrogen gas will similarly decrease as gas clouds condense into stars and the hydrogen itself is 'burnt' by the component stars. By this method, the age of star clusters and an estimate of the age of a galaxy can also be determined.

There is a limit to this process. Once silicon has fused via the alpha process and iron has been produced, no further fusion reactions will occur as they would require a net input of energy. The star begins gravitational collapse again. As they do, more heating occurs and, as a result, stars up to about $8M_\odot$ tend to shed some of their mass by ejecting what are called **planetary nebulae** (see Figure 13.3.8). These are basically layers formed as a result of the various stages of fusion reactions and may involve a considerable proportion of the star's mass. Inside there remains a small hot star. Deprived of a source of heat, such as thermonuclear reactions, the star will cool down for thousands of millions of years to finally become a black dwarf.

FIGURE 13.3.8 The Helix nebula is classed as a planetary nebula and lies around 700 light-years from Earth in the constellation Aquarius.

Neutron stars and pulsars

For the true giants, over $8M_\odot$, their very high gravity doesn't allow the ejection of planetary nebula and so they continue to collapse, generating yet more heat. And then in an amazing sequence of events not totally understood, the core, largely of iron by this stage, rapidly skyrockets to many billions of degrees in a fraction of a second. This results in the protons and electrons in the core being forced together with such ferocity that they combine to form neutrons and small almost-massless particles called neutrinos.

The core by this stage is about 20 kilometres in diameter but with a mass many times that of the Sun. In this state the core can collapse no further, but the rest of the mass of the star is still collapsing in on it, building up enormous pressure which, as the core can compress no further, creates an absolutely monumental 'bounce back'. This sends the outer layers of the star flying off into space at enormous speeds, partly propelled by the neutrinos trying to escape from the core. The energy released in this explosion is an incredible 10^{46} J, which is far more than the Sun will produce in its entire life.

FIGURE 13.3.9 The Crab Nebula lies around 6000 light-years from Earth in the constellation Taurus. In the centre of the nebula is the remnant of the star's core, which has formed a pulsar, rotating 30 times a second and emitting regular pulses of X-ray radiation.

What is left of the star after this gigantic explosion is an extremely dense core made of 'neutron matter'. This is matter in which all the electrons and protons have collapsed to form neutrons. A **neutron star** with a diameter of 30 kilometres would have the same mass as the Sun. A few cubic centimetres would have a mass of hundreds of millions of tonnes.

In the 1960s a number of **pulsars** were found. These were objects producing extremely regular pulses of radio emissions. Later, it was found that these pulses could also be seen as flashes in visible light. They are now thought to emanate from rotating neutron stars about 20 kilometres in diameter but with about 1.4 solar masses. The frequency of rotation of a star is greatly increased as it collapses, in just the same way that skaters or dancers can spin faster by pulling their arms closer to their body as they spin.

The **periods** of pulsars are measured in seconds or even fractions of a second—they are spinning at an incredible rate for their size. As they spin, leftover charged matter interacts with their powerful magnetic fields and produces a beam of intense radiation that sweeps around the sky like the beam from a lighthouse (Figure 13.3.9). If observers happen to be in line with the beam, a flash is seen each time it sweeps past.

The other remains from supernovae are nebulae (dust and gas clouds). An example of such a nebula is the Crab Nebula (seen in Figure 13.3.9 and 13.3.10). The Crab Nebula is the remnant of a supernova that was first seen from Earth in 1054. Nebulae are the ejected gases that glow for possibly hundreds of years after the supernova. Gradually they fade and disperse into space to form more of the dust and gas clouds out of which new stars will be born. Many of these remnants can only be seen by radio telescopes, as their temperature drops and the wavelength of the radiation becomes longer than that of visible light.

FIGURE 13.3.10 An image from the Hubble Space Telescope of the Crab nebula, denoted M1. The nebula is the remnants of a supernova that was visible from Earth in 1054 and was bright enough to be seen during the day.

Black holes

Apart from being a marvellous spectacle, supernovae are vital to the existence of Earth and humankind. The processes in large stars can produce all the elements up to iron (atomic number 26), because the fusion of nuclei up to iron result in a net release of energy. However, to fuse heavier nuclei requires massive amounts of energy. You may remember the binding energy curve from Chapter 7 (p. 236), which explained this.

The only situation in which sufficient energy is produced to create these larger nuclei is a supernova. Among the material thrown into space by a supernova are all the elements of the periodic table. Billions of years later they may become part of a new planetary system, a so-called second-generation system, which will include all the elements needed for life to emerge. Our Sun and its planets are such a system. Many of the atoms in your bodies were initially created in a supernova.

Even more fascinating is what happens to really massive stars, say over $20M_{\odot}$, once they run out of fuel. There is a point at which even the neutrons in a super-heavy neutron star will collapse. It seems that this further collapse has no end point at all and matter simply collapses into what is called a **black hole**. More particularly it is referred to as a 'singularity'—a point of infinitely small volume into which matter is squeezed. The force of gravity around such an object is so strong that anything, including light, that comes close enough is just 'sucked' in.

Figure 13.3.11 shows the Perseus constellation, which contains the galaxy NGC1277. At the centre of this relatively small galaxy is one of the most massive black holes ever found, with a mass equivalent to 17 billion Suns, and which makes up more than 14% of the total mass of the galaxy. This compares with the one thought to be at the centre of the Milky Way, making up just 0.1% of the Milky Way's mass. How such a massive black hole came to form in the centre of a galaxy is causing scientists to question their theories on how black holes form.

FIGURE 13.3.11 An image from the Hubble Space Telescope of the Perseus constellation.

Around a black hole is the **event horizon**, an imaginary spherical boundary that marks the 'point of no return'. At the event horizon, it is impossible for even light to escape the gravitational pull of the black hole. The radius of this spherical boundary was determined by a German physicist Karl Schwarzschild and is called the **Schwarzschild radius** in his honour.

The Schwarzschild radius can be found by using the equation:

$$r_s = \frac{2GM}{c^2}$$

where r_s is the Schwarzschild radius in metres (m)

G is the universal gravitational force constant, 6.7×10^{-11} N m^2 kg^{-2}

M is the mass of the collapsed neutron star in kilograms (kg)

c is the speed of light, 3.0×10^8 m s^{-1}.

The fact that c^2 is so large, and G is very small, is an indication that the mass of the original star has to be extremely large if a black hole is going to form.

Astrophysicists can't actually see black holes but evidence of their existence can be gained from the study of the intense gravitational effects that they have on the objects around them. A black hole passing through a cloud of interstellar matter will draw matter towards it in a process known as **accretion**. A normal star passing close to a black hole will be torn apart, emitting bursts of X-rays that radiate into space as it falls into the black hole. Black holes can also trigger the growth of new stars as their gravity influences the clouds of dust and gas around them.

FIGURE 13.3.12 Black holes draw matter towards them, forming a vast accretion disc—a donut-shaped gas cloud.

Worked example 13.3.2

SCHWARZSCHILD RADIUS

Were it possible for the Sun to collapse into a black hole, what would its Schwarzschild radius be in kilometres?
(Use $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}, $M_\odot = 2.0 \times 10^{30}$ kg and $c = 3.0 \times 10^8$ m s^{-1}.)

Thinking	Working
Use the equation for calculating the Schwarzschild radius.	$r_s = \frac{2GM}{c^2}$
Substitute the known values into the equation for the Schwarzschild radius and solve for r_s.	$r_s = \frac{2GM}{c^2}$ $= \frac{2 \times 6.7 \times 10^{-11} \times 2 \times 10^{30}}{(3 \times 10^8)^2}$ $= 2978$ m $= 3.0$ km

Worked example: Try yourself 13.3.2

SCHWARZSCHILD RADIUS

A star 30 times the mass of the Sun collapses into a black hole. What would its Schwarzschild radius be in kilometres?
(Use $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}, $M_\odot = 2.0 \times 10^{30}$ kg and $c = 3.0 \times 10^8$ m s^{-1}.)

Black holes, space and time

Black holes are the most compact form of matter known. When a star collapses, its whole mass is concentrated into a fraction of the space of the original star. While it hasn't suddenly gained more mass, the effect of concentrating the mass in such a small area creates a gravitational field that does some very strange things to space and time.

Einstein never really believed that black holes existed. However, his theories allowed for their existence. In developing his theory of general relativity, Einstein combined the three dimensions of physical space and time into one idea that he referred to as **spacetime**.

PHYSICSFILE

The multiverse

Recent work attempts to explain singularities and more with the idea of a multiverse. A multiverse refers to multiple parallel universes, each separate from the other. To create a new universe you would need to start with a massively dense, infinitely small singularity cut off from the rest of the universe in which it exists. This is the theorised state of our universe just before the big bang. Humankind may owe its existence to a black hole in some other part of the multiverse.

In spacetime, all four dimensions are interdependent. Motion through space comes at the expense of motion through time. There is no universal time. Time will go at different rates for people travelling at different relative speeds. If acceleration is equivalent to gravitation (recall the acceleration due to gravity studied in motion) then this means that time is affected by gravity. GPS systems actually use this effect to our advantage. The clocks on GPS satellites run very slightly slower than those on Earth and that difference is used in calculating their positions relative to a point on Earth. This time difference is called gravitational time dilation.

In 'normal' space, the effect of gravitational time dilation is only just measurable. Around a black hole, gravitational effects are extreme given the enormous concentration of mass. The effect is to actually warp or curve spacetime. The distance a beam of light has to cover within the gravitational field of a black hole to get from one spot to another is significantly larger and hence takes longer.

The strange part is, an observer will see an object travelling slower and slower as it approaches the event horizon of a black hole. When the object is right near the event horizon, its motion is so slow that it would not actually reach it and would appear frozen at the black hole's edge. From the object's perspective time moves on at the same rate it always has—that is, until it's ripped apart by the gravitational field within the black hole.

GALAXIES

Galaxies are massive collections of stars that are bound together by gravitational attraction. Studies by the Hubble Space Telescope and Earth-based observatories suggest that there may be at least as many galaxies in the universe as there are stars in the Milky Way Galaxy (see Figure 13.3.13). The Hubble Telescope viewed just one small patch of sky around one-tenth the diameter of the Moon for less than 12 days and found around 10 000 galaxies of varying size, shape and colour.

FIGURE 13.3.13 A cluster of galaxies dominates the field of view of what at first glance appears to be just a field of stars. In fact, most of the light sources in this photo are galaxies.

The study of galaxies also provides information about the life and evolution of the stars within and poses new questions about how galaxies form. Much has been learnt, but the fundamental theories on exactly how galaxies form and evolve are still to be confirmed.

Forming galaxies

Scientists originally thought that matter was evenly spread throughout the universe, but a study started in 2003 made them rethink this.

The original big bang that is theorised to have created the universe should have left some radiation behind as a remnant—cosmic microwave background radiation, as discussed in Chapter 5. The Wilkinson Microwave Anisotropy Probe (WMAP) spacecraft was launched to study its distribution and to confirm its existence. The completed sky survey confirmed the existence of this background radiation, providing further firm evidence for the big bang. It was also able to establish from the wavelengths of the radiation that the universe is 13.7 billion years old and is expanding at an astonishing rate. It also revealed that matter was not evenly spread throughout the universe.

Tiny quantum fluctuations in the time immediately following the big bang created regions of higher and lower densities. In Figure 13.3.14 on page 483, higher density regions are coloured red and yellow. Regions of high density formed the seeds of early galaxies. These early galaxies consisted mostly of gas. As gravity compressed the gas, stars started to form, eventually creating the galaxies as we know them today.

PHYSICSFILE

Dark matter in spiral galaxies

Astronomers have a number of theories about how spiral galaxies are formed. The most recent theories use the idea that dark matter (see Chapter 5, p. 175) forms halos around the outer parts of the galaxy. Dark matter is thought to only interact gravitationally and so doesn't disperse, but rather acts to form a thin, rapidly rotating disc as the gas and dust that makes up the visible early galaxy contracts. None of the proposed theories has yet completely explained the spiral galaxies' size and rotation.

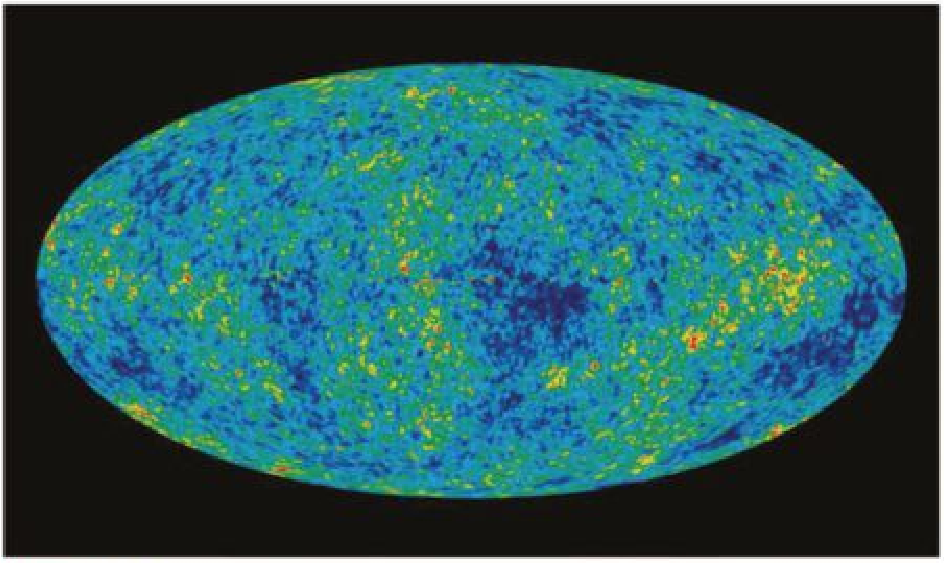


FIGURE 13.3.14 The cosmic microwave background radiation of the whole sky. Variations in colour reveal the different densities of the early universe. Denser regions, shown in red and yellow, became the early galaxies and other structures that make up the universe today.

Size and shape

Earth's own galaxy is a form of disc or spiral galaxy. These galaxies are relatively thin and rotate rapidly, at least on a cosmic time scale. They are only one of many forms of galaxy (see Figure 13.3.15) ranging from large, almost featureless collections of old stars called elliptical galaxies through to various forms of spiral and lenticular galaxies.

The Milky Way is classified as a barred spiral galaxy with well-defined arms, although this is not certain as there is no way for an observer to get a good look from the outside. There are also galaxies that don't fit into any of these categories, termed 'irregular' galaxies. The Greater and Smaller Magellanic Clouds, which are both satellite galaxies of the Milky Way, are examples of dwarf irregular galaxies.

Colour

The spectra of distant galaxies are shifted to the red end of the spectrum due to their movement away from Earth as the universe expands. This is only one feature of a galaxy that influences its colour.

The largest galaxies are elliptical galaxies. They are smooth and relatively featureless and are usually reddish in colour, indicating that they are composed largely of older stars and little or no star-building dust and gas. Most elliptical galaxies investigated also have supermassive black holes at their centres and are found in more-crowded regions of the universe. Astronomers view them as the most evolved of the galaxies—older and with few if any new stars being created. The emission spectra of these older galaxies will typically also show more of the heavier elements created when supernovas explode.

Elliptical galaxies are thought to have been created from the collision of smaller galaxies because galaxies tend to be found in clusters bound by gravity to each other. The Milky Way is currently drawing in the Magellanic Clouds and is ripping apart a very small satellite galaxy called Sagittarius Dwarf Elliptical Galaxy. It is also bound by gravity to the vast Andromeda (M31) spiral galaxy.

In the future, the Milky Way and Andromeda galaxies will merge at high speed (speeds of up to 500 kilometres per second have been observed in other merging galaxies) with gravity distorting and remodelling both galaxies until eventually they form a single large elliptical galaxy. Andromeda is already partly distorted from what are thought to have been mergers with smaller satellite galaxies in the past. These mergers also generate new dust and gas, creating new star-forming regions. The image in Figure 13.3.16 (p. 484) shows how this sort of merger might happen.

FIGURE 13.3.15 Fours types of galaxies. From top to bottom: lenticular, spiral, irregular and elliptical.

FIGURE 13.3.16 A supercomputer image of two spiral galaxies colliding. The Milky Way and Andromeda will go through a similar process in the next few billion years.

Spiral galaxies make up the largest number of galaxies in today's universe. Typically they are blue-coloured due their younger, brighter stars and new star formation that is still happening within them.

Similar in concept to the H–R diagram for stars, a galaxy colour–magnitude diagram has been developed showing the relationship between the mass and the absolute magnitude of galaxies (see Figure 13.3.17). The diagram includes three main groupings:

- red sequence—most red and largely elliptical galaxies
- blue cloud—most galaxies and generally spirals, including galaxies in the process of large-scale star formation
- green valley—some galaxies later in their evolution, including red spirals, where star formation has stopped. Star formation stops when the gas to form stars has run out, where the gas has been destroyed very quickly due to mergers, or where gas has been lost to what are believed to be supermassive black holes at their centres.

Both the Milky Way and the Andromeda galaxies are thought to lie in the green valley. Large-scale star formation is decreasing as the supply of star-forming gas declines.

Unlike stars, the properties of galaxies are not completely determined by their colour and magnitude.

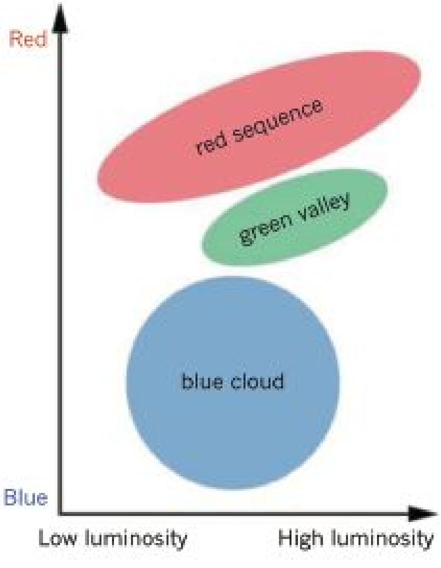


FIGURE 13.3.17 The galaxy colour–magnitude diagram shows the relationship between mass and absolute magnitude for galaxies.

13.3 Review

SUMMARY

- The Sun produces energy by the fusion of hydrogen nuclei to form helium nuclei.
- Calculations based on the mass and energy output of the Sun showed that the amount of energy being produced for each atom in the Sun was around a hundred million times greater than the energy produced by each atom in chemical reactions. This can be explained by mass–energy equivalence and can be found by Einstein's equation $E = mc^2$.
- The light we see from the Sun is emitted from the photosphere, the temperature of which is approximately 5800 K.
- Most stars begin their lives on the main sequence of the H–R diagram, eventually become giants and then move downwards to become dwarf stars.
- As stars age, the products of nuclear fusion within the star change. Young stars fuse hydrogen to form helium. As stars age they produce increasingly heavy elements. Once silicon starts to be fused to produce iron, further nuclear fusion reactions would require a net input of energy and thus do not occur. The star begins to collapse under gravitational forces.
- Some very large stars end their lives as supernovae, which create the heavier elements, allowing the creation of new stars, planets and life itself.
- The age of a star, star clusters and the average age of galaxies can be determined by the change in chemical spectra as successive fusion processes occur in the component stars.
- Some massive stars collapse to form super-dense neutron stars, pulsars and black holes.
- Around a black hole is the 'event horizon', an imaginary spherical boundary at which it is impossible for even light to escape.
- The radius of the event horizon for a particular black hole is called the Schwarzschild radius and can be found by: $\frac{2GM}{c^2}$
- Einstein combined the three dimensions of physical space and movement in time into one idea that he referred to as spacetime. In spacetime, all four dimensions are interdependent.
- In 'normal' space, the effect of gravitational time dilation is only just measurable. Around a black hole, gravitational effects are extreme given the enormous concentration of mass. The effect is to actually warp or curve spacetime.
- An observer would see light in a black hole slow down, theoretically to the point where an object falling past the event horizon of a black hole would appear frozen at the black hole's edge. From the object's perspective, time will move on at the same rate it always has.
- Scientists originally thought that matter was evenly spread throughout the universe. The WMAP sky survey established that there are high and low density regions. Regions of high density formed the seeds of early galaxies.
- Galaxies can be categorised by size, shape and colour.
- There are four main types of galaxies: spiral, elliptical, lenticular and irregular.
- The largest galaxies are elliptical galaxies. They are smooth and relatively featureless and are usually reddish in colour, indicating that they are composed largely of older stars and little or no star-building dust and gas. They are thought to form from the merger of spiral galaxies.
- Spiral galaxies make up the largest number of galaxies in today's universe. Typically they are blue-coloured due the younger, brighter stars and new star formation that is still happening within them.
- A galaxy colour–magnitude diagram has been developed showing the relationship between the mass and the absolute magnitude of galaxies. Three main groupings are included—red sequence (old, elliptical galaxies), blue cloud (star-producing spiral galaxies) and green valley (galaxies such as the Milky Way where star production has slowed). Unlike stars, the properties of galaxies are not completely determined by their colour and magnitude.

13.3 Review *continued*

KEY QUESTIONS

1 Why is it that the nuclear reactions involving hydrogen in the Sun can produce so much more energy than hydrogen burning in oxygen on the Earth?

2 Which layer of the Sun's atmosphere gives rise to the sunlight that arrives on Earth?

A chromosphere
B corona
C photosphere
D all layers

3 A star is losing mass at the rate of 6×10^9 kg s^{-1}. How much energy is being produced per second in total radiation if this is the case? (Use $c = 3.0 \times 10^8$ m s^{-1}.)

4 What is the order of stellar evolution for most stars?

5 Near the end of their lives, some stars, with masses near $8M_\odot$, shed some of their mass as a series of shells of gas. What are these shells of gas called?

6 The most likely scenario for the end of the Sun once it has exhausted its fuel supply is to finish as what kind of astronomical object?

7 Around a black hole is the event horizon, a distance from the black hole at which it is impossible for even light to escape. What is this distance also referred to as?

8 A star 25 times the mass of the Sun collapses into a black hole. What would its Schwarzschild radius be in kilometres (to two significant figures)? (Use $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}, $M_\odot = 2.0 \times 10^{30}$ kg and $c = 3.0 \times 10^8$ m s^{-1}.)

9 If it were possible for Jupiter to collapse into a black hole, what would be the radius of its event horizon? (Use $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}, $M_{\text{Jupiter}} = 1.9 \times 10^{27}$ kg and $c = 3.0 \times 10^8$ m s^{-1} and give your answer to two significant figures).)

10 Galaxies are grouped on the galaxy colour–magnitude diagram. This is not as definitive as the H–R diagram is for stars but does show the relationship between mass and absolute magnitude. Determine the type of galaxy based on star production with each of the main sequences shown in the galaxy colour–magnitude diagram for:

a no star production
b large-scale production
c slowing star production.

Chapter review

13

KEY TERMS

absolute brightness
absolute magnitude
accretion
apparent brightness
apparent magnitude
binary stars
black dwarf
black hole
diffract
diffraction grating
event horizon
frequency
giant
Hertzsprung–Russell diagram
hydrostatic equilibrium
interference pattern
intrinsic brightness
luminosity
main sequence
neutron star
period
photosphere
planetary nebula
protostar
pulsar
refraction
Schwarzschild radius
spacetime
spectroscope
spectroscopic parallax
standard candle
stellar parallax
supernova
white dwarfs

For any of the questions that follow, use $c = 3.0 \times 10^8$ m s^{-1}, $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}, unless otherwise specified.

The following information refers to questions 1 and 2.

Cosmic microwave background radiation provides evidence supporting the big bang theory. It has a peak wavelength of 1.869 mm, corresponding to a temperature of $T = 2.762$ K.

1 What is the corresponding frequency of this radiation?

2 What is the corresponding period for the frequency of the cosmic microwave background radiation?

3 The distance to some close stars can be measured via their apparent movement relative to the position of the Earth as it orbits the Sun. What is this method for determining a star's position referred to as?

A absolute magnitude
B stellar parallax
C standard candle
D red shift

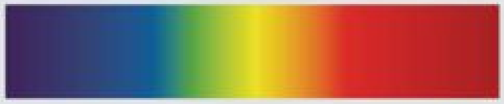
4 Sirius is the brightest star in the sky when viewed from Earth. Canopus is the second brightest when viewed from Earth but is actually 3000 times brighter than Sirius. How can this be?

5 Rigel and Deneb are two stars with the same absolute magnitude: –7.1. Rigel has an apparent magnitude of 0.14 and is much closer to the Earth than Deneb. Based on this information which of the following would be a reasonable estimate for the apparent magnitude of Deneb?

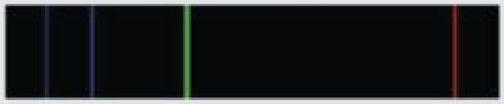
A –7.1
B –1.26
C 0.14
D 1.26

6 Label each of the photos of the spectra shown with the type of spectrum.

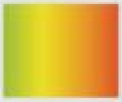
a



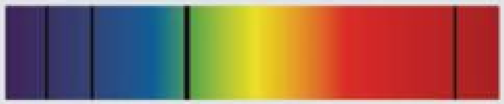
b



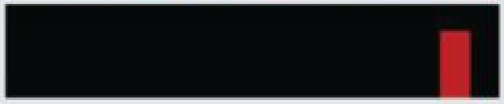
c



d



e



7 Which characteristics of a star can you directly determine from its spectrum?

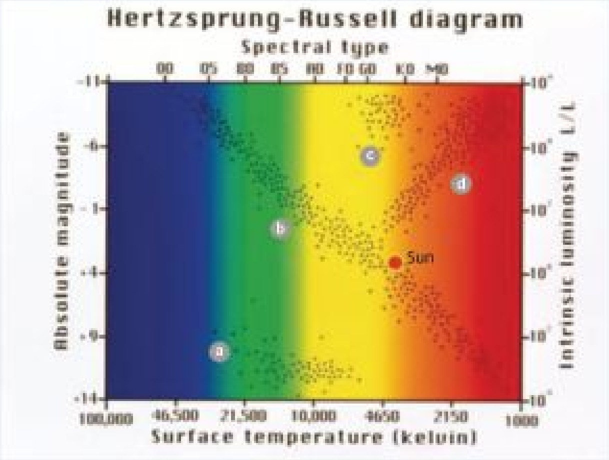
8 As the stars are so far away, how can scientists be so sure that they are not made of totally new elements never seen before on Earth?

9 Referring to the Figure 13.2.5 on page 469, order the following stars from smallest to largest.

Tau Ceti
Aldebaran
Rigel
Vega
Barnard's Star

Chapter review *continued*

10 Below is a representation of the Hertzsprung–Russell (H–R) diagram in which stars are grouped by luminosity and temperature. Label the diagram with the types of stars that would be found in each part of the diagram.



11 Rigel is a blue supergiant star located in the middle-top of the H–R diagram. Based on this information describe briefly, in point form, the main stages of its expected life cycle.

12 In terms of the H–R diagram, where are stars 'born'?

13 The Sun is an 'average' star. In what sense is this statement true in regard to the H–R diagram, and in what sense is it not true?

14 Delta Orionis is the seventh brightest star in the constellation Orion. It has a mass 20 times that of the Sun and a surface temperature of 20 000 K. If it were to collapse into a black hole, what would be the Schwarzschild radius? (Use $M_{\odot} = 1.989 \times 10^{30}$ kg.)

15 What is another name for a black hole?

16 To an observer, how does light appear at the event horizon of a black hole?

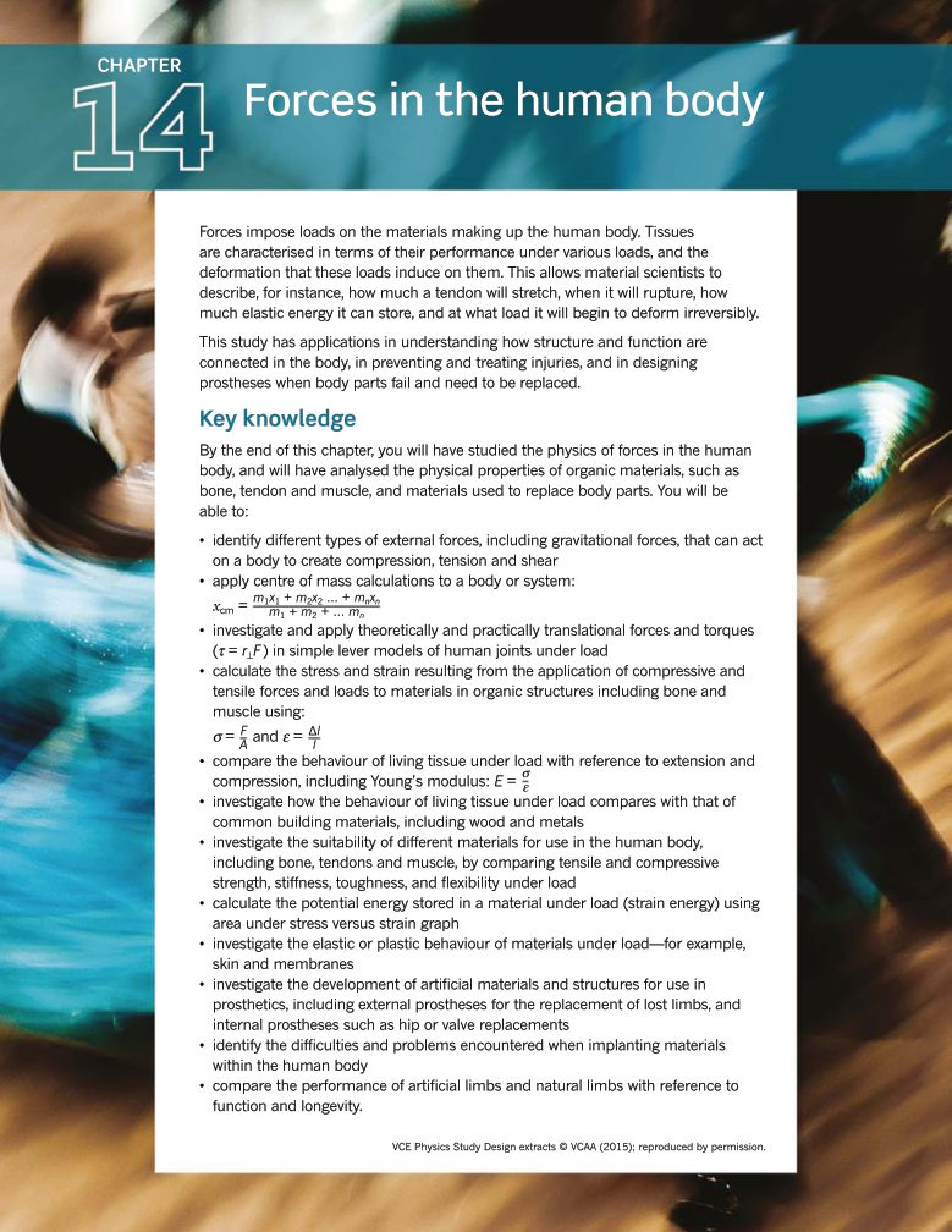
17 A very large elliptical galaxy has been discovered. The peak wavelength of the light from the galaxy lies in the red region of the visible spectrum. What assumption(s) can be reasonably made about this galaxy based solely on this information?

A Star production will be low or will have ceased.
B A supermassive back hole will lie at its centre.
C It is the product of one or more merges of smaller galaxies.
D It is the result of a supernova explosion.
E New star formation is happening at the centre at an even rate.

18 A star is losing mass at the rate of 4×10^9 kg s^{-1}. How much energy is being produced per second in total radiation?

19 If it were possible for Earth to collapse into a black hole, what would be its Schwarzschild radius (to two significant figures)? Use $M_{\text{Earth}} = 6.0 \times 10^{24}$ kg.

20 A star collapses into a black hole that has a Schwarzschild radius of 100 km. Determine how many times heavier than the Sun this star must be. Use $M_{\odot} = 2.0 \times 10^{30}$ kg.



CHAPTER

14 Forces in the human body

Forces impose loads on the materials making up the human body. Tissues are characterised in terms of their performance under various loads, and the deformation that these loads induce on them. This allows material scientists to describe, for instance, how much a tendon will stretch, when it will rupture, how much elastic energy it can store, and at what load it will begin to deform irreversibly.

This study has applications in understanding how structure and function are connected in the body, in preventing and treating injuries, and in designing prostheses when body parts fail and need to be replaced.

Key knowledge

By the end of this chapter, you will have studied the physics of forces in the human body, and will have analysed the physical properties of organic materials, such as bone, tendon and muscle, and materials used to replace body parts. You will be able to:

- identify different types of external forces, including gravitational forces, that can act on a body to create compression, tension and shear
- apply centre of mass calculations to a body or system: $x_{cm} = \frac{m_1x_1 + m_2x_2 \ldots + m_nx_n}{m_1 + m_2 + \ldots m_n}$
- investigate and apply theoretically and practically translational forces and torques ($\tau = r_\perp F$) in simple lever models of human joints under load
- calculate the stress and strain resulting from the application of compressive and tensile forces and loads to materials in organic structures including bone and muscle using: $\sigma = \frac{F}{A}$ and $\varepsilon = \frac{\Delta l}{l}$
- compare the behaviour of living tissue under load with reference to extension and compression, including Young's modulus: $E = \frac{\sigma}{\varepsilon}$
- investigate how the behaviour of living tissue under load compares with that of common building materials, including wood and metals
- investigate the suitability of different materials for use in the human body, including bone, tendons and muscle, by comparing tensile and compressive strength, stiffness, toughness, and flexibility under load
- calculate the potential energy stored in a material under load (strain energy) using area under stress versus strain graph
- investigate the elastic or plastic behaviour of materials under load—for example, skin and membranes
- investigate the development of artificial materials and structures for use in prosthetics, including external prostheses for the replacement of lost limbs, and internal prostheses such as hip or valve replacements
- identify the difficulties and problems encountered when implanting materials within the human body
- compare the performance of artificial limbs and natural limbs with reference to function and longevity.

14.1 Forces acting on the body

FIGURE 14.1.1 Researchers use crash test dummies to simulate the effects of high-speed car crashes in order to understand the impacts on the human body.

Researchers simulate crashes using a dummy to find out what would happen to a person in a high-speed car accident (Figure 14.1.1). You may already know that a passenger continues to move forwards when a car hits a lamppost. The passenger's torso is restrained by the seatbelt, but the head is free to lurch forwards. This stretches and potentially overextends muscles, ligaments and tendons in the neck, creating an injury commonly known as whiplash. If the skull hits the windscreen, it could sustain a fracture, and a concussion could occur. A concussion occurs when the soft tissue of the brain continues to move forwards while the skull stops. Body tissues can be stretched, compressed, twisted or subjected to shearing action. In these stresses, the specific physical properties of the different biological materials and the magnitudes and directions of the forces acting affect the injury.

FORCES ON THE BODY: CONTACT FORCES AND GRAVITY

External forces may be applied to the human body as a result of acceleration, such as the force of the aeroplane seat on your back as you accelerate to take off, or deceleration, such as the force on your palm as you catch a cricket ball in the outfield. Recall that these are called contact forces since they act between objects in physical contact with each other. Contact forces are in contrast to forces, such as electrostatic forces and gravity, that can act at a distance. Forces that can act at a distance are called non-contact forces.

Because the human body has mass, it experiences the **gravitational force** of the Earth, which is commonly called weight. This force attracts the body towards the centre of the Earth. As the body presses against a surface, it experiences the **normal reaction force** pushing upwards. A body stationary on Earth experiences weight and normal reaction forces that are equal in magnitude but in opposite directions. These forces are shown in Figure 14.1.2.

Forces are usually described by the notation $F_{\text{on A by B}}$, which means the force on A by B. So, for example, the weight of the man in Figure 14.1.2 is the gravitational force on the man by the Earth, and would be denoted $F_{\text{on M by E}}$, while the force with which the man attracts the Earth would be $F_{\text{on E by M}}$.

External forces give rise to internal forces in the body. The combined effect of gravity and the normal reaction force compress the load-bearing structures of the human body. The normal reaction force is applied at the feet and transmitted upwards, while load-bearing structures experience a downwards force from the weight of the body above them. The result is compression of the leg bones, pelvis, vertebrae and the discs of the spinal column.

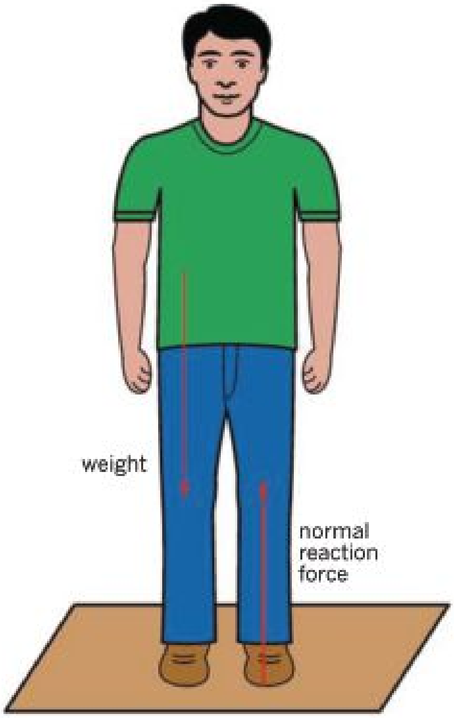


FIGURE 14.1.2 A person's weight is the gravitational force on the person applied by the Earth, and the normal reaction force is the force on the person applied by the Earth's surface. These forces are equal in magnitude and opposite in direction when the person is at rest.

Astronauts on the Moon experience a reduced gravitational force and reaction force, hence the compressive forces on their load-bearing body parts are less. Astronauts in orbit around the Earth on the International Space Station are in a microgravity environment. This has consequences for their bones, muscles and tendons, which tend to deteriorate under conditions of zero load. For this reason astronauts need to exercise diligently in order to maintain physical condition.

FORCES CAUSE ACCELERATION

Recall from Chapter 11 that when the forces on an object are balanced, the body is in translational equilibrium. An unbalanced force will cause acceleration of body parts, or of the whole body. Alternatively, forces can also cause rotation about an axis, which is evident as bones pivot about the joints. A body that is neither accelerating nor subject to a torque is said to be in translational *and* rotational equilibrium, implying that the vector sum of the forces and the vector sum of the torques acting on the body are both zero.

Translation

Movement in the human body requires the interaction of muscles, bones and tendons. The muscles connect to bones via tendons, as shown in Figure 14.1.3. When muscles contract, they pull on the tendons, which then apply forces to the bones, and this causes movement, or **translation**. The complex movements of body parts caused by muscle contraction, in addition to the action of the gravitational force and external forces such as friction, allow the body to perform complex movements such as dancing or surfing.

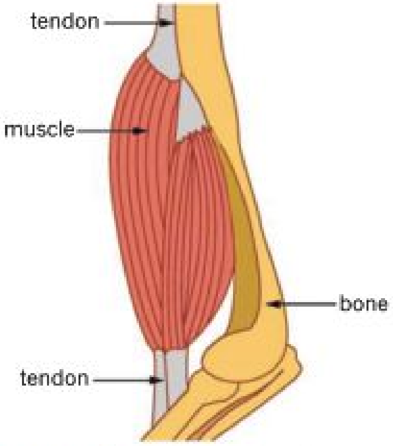


FIGURE 14.1.3 This diagram shows how muscles and tendons connect the bones of the upper and lower arm across the elbow joint. Tendons attach muscles to bones.

FORCES CAUSE INTERNAL STRESS

Forces applied at two or more points on a body may not cause movement, but cause internal stresses on the tissues. These can bring about deformation, and in extreme cases even failure of the body structures.

Compressive forces

When a material is being squashed or squeezed, the material is said to be under compression. This is caused by forces which are applied in opposing directions towards each other, as shown in Figure 14.1.4. **Compressive forces** push the atoms in the material closer together.

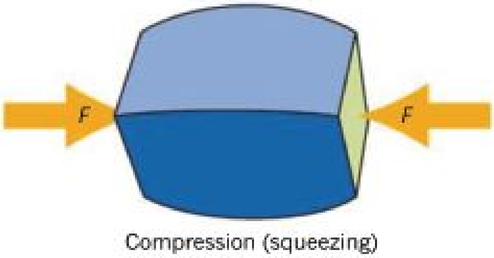


FIGURE 14.1.4 Internal stresses: opposing forces directed towards each other cause compression.

When the weight of the body above the knee exerts a force on the bones of the lower leg, and the normal reaction force is transmitted through the bones of the feet, the tibia (shin bone) experiences compressive forces. The diagram in Figure 14.1.5 shows how these compressive forces act on the tibia. The forces are shown as blue arrows pointing towards each other.

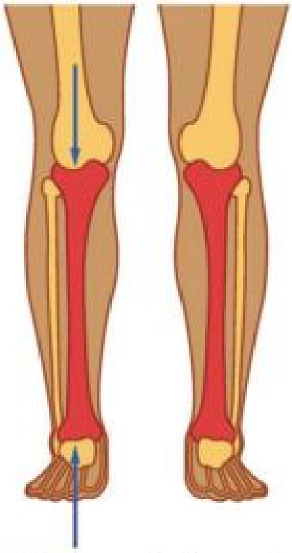


FIGURE 14.1.5 Compressive forces acting on the tibia.

The ability of bones to bear a load depends on the compressive stress that they routinely experience. People who do a lot of load-bearing exercise have higher bone density, and their bones become stronger than those of people who do not. Conversely, astronauts who spend prolonged time in microgravity environments, where there is little load-bearing, tend to lose bone density as well as muscle and tendon condition.

Tensile forces

When a stretching force acts on a material, the material is said to be under tension. This is caused by forces that are exerted in opposite directions away from each other, as shown in Figure 14.1.6. **Tensile forces** pull the atoms in the material further apart.

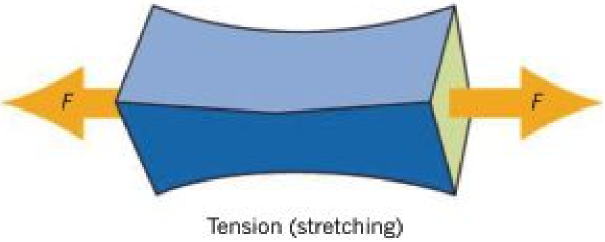


FIGURE 14.1.6 Internal stresses: opposing forces directed away from each other cause tension.

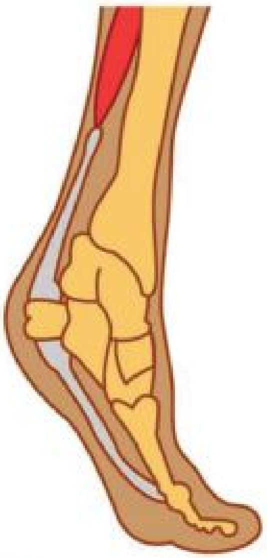


FIGURE 14.1.7 The Achilles tendon attaches the calf muscle to the heel.

One place where tension forces act in the body is through the Achilles tendon, at the back of the ankle. The Achilles tendon attaches the calf muscle to the heel bone (see Figure 14.1.7). When a person crouches down but keeps all of their weight on their toes, their Achilles tendon is under tension. Another way to put the tendon under tension is to stand on the edge of a step, with the heel dropped down over the edge. In this position, the heel pulls downwards on the tendon. The result is that the tendon is stretched.

Shear forces

If a force acts across the top of an object while the bottom remains fixed, a shearing effect or **shear force** is said to act. The easiest way to imagine this is to think of a thick book on a table. Friction keeps the bottom of the book from moving, but the book will shear if a force is applied towards the spine on the top of the book, as shown in Figure 14.1.8. The object experiencing the shear force does not change significantly in its dimensions, but it does change shape.

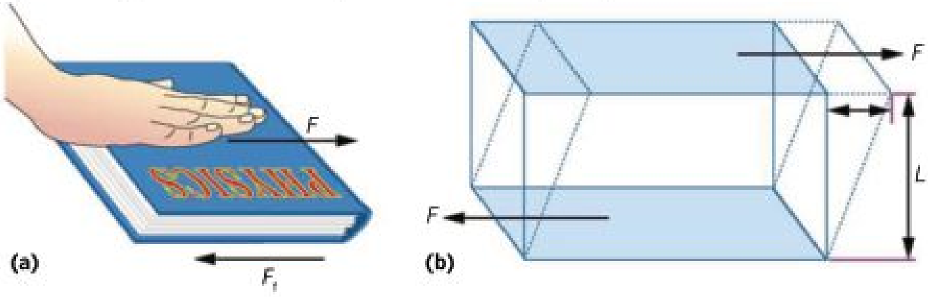


FIGURE 14.1.8 (a) As the book is pushed while remaining at rest, an equal and opposite force is supplied by the bench. The resulting shear causes the book to change shape. (b) This diagram shows the effect of the forces on the shape of the book.

Slide tackling in soccer is a dangerous manoeuvre. Typically the player in possession of the ball has his or her feet firmly planted on the ground where weight and frictional forces secure the lower part of the leg. In side tackling, an opponent comes sliding in with high momentum, hoping to make contact with the ball. Should the opponent misjudge and contact the player somewhere on the leg, enormous shearing forces are generated. These shear forces could fracture bones (see Figure 14.1.9).

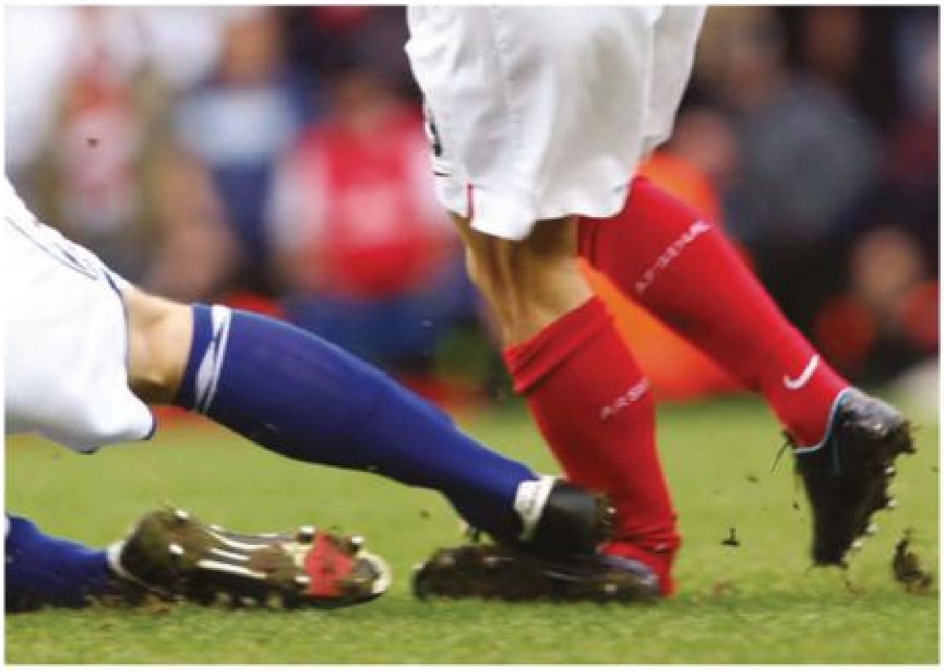
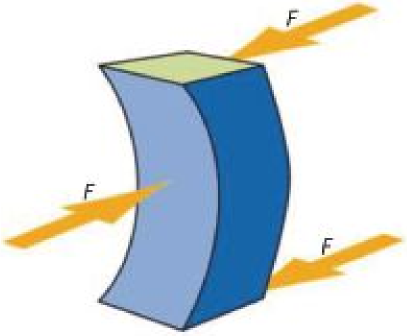


FIGURE 14.1.9 This slide tackle caused shearing forces that have fractured both bones of the lower leg.

Bending

The application of multiple forces on an object can result in bending, as shown in Figure 14.1.10.



Bending

FIGURE 14.1.10 Internal stresses: bending causes a combination of tension and compression in a material. Here, tension is experienced at the back of the object (on the convex curve) and compression at the front (on the concave curve).

When an object bends, it experiences compression along the face where the push of the force acts. The other side of the object will experience tension. There is also an element of shear when objects bend. Loading of a long bone when the ends are not free to move will cause the bone to bend. As bone is stronger under compression than tension, any failure will occur on the side that is under tension, as shown in Figure 14.1.11.

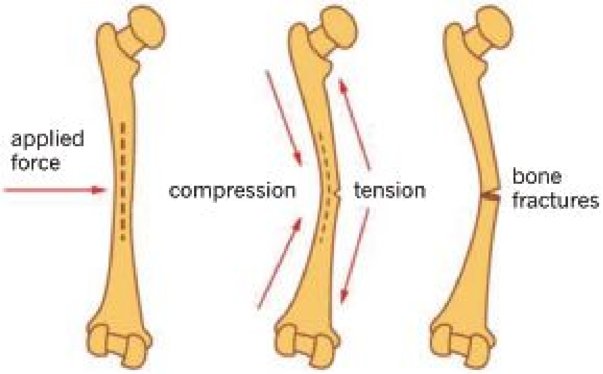


FIGURE 14.1.11 A force applied in the middle of a long bone causes bending. The concave side of the bone, where the force is applied, is under compression, while the convex side is under tension. The bone will fracture if the tension is too great.

This type of fracture is particularly common in children because their bones are softer and more flexible. The bone only breaks partially, resulting in a so-called greenstick fracture. Adults are more likely to experience a clean break because their bones are more brittle.

CENTRE OF MASS

When studying forces and motion, rigid bodies are considered as having a mass concentrated at the centre. The human body does not move in the way a more symmetrical object would. It is often helpful to treat the body as a collection of separate masses. Each part then has its own centre, where the mass appears to be concentrated.

Finding the centre of mass

The balancing point of an object is called the centre of mass. It is the point at which the mass appears to be concentrated, and so it is the point that moves as though all the net external forces acted on it.

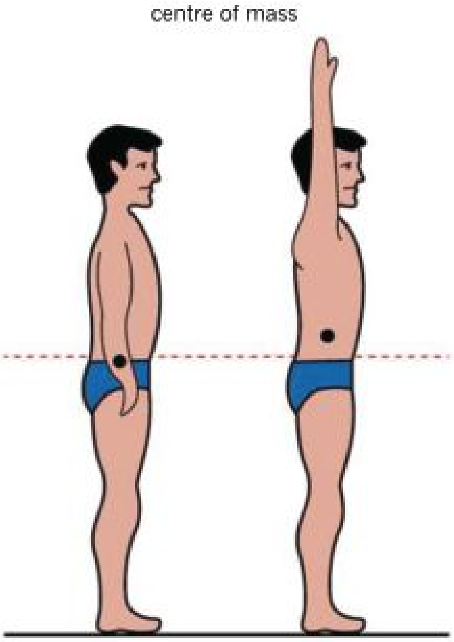


FIGURE 14.1.12 The centre of mass moves as body position changes.

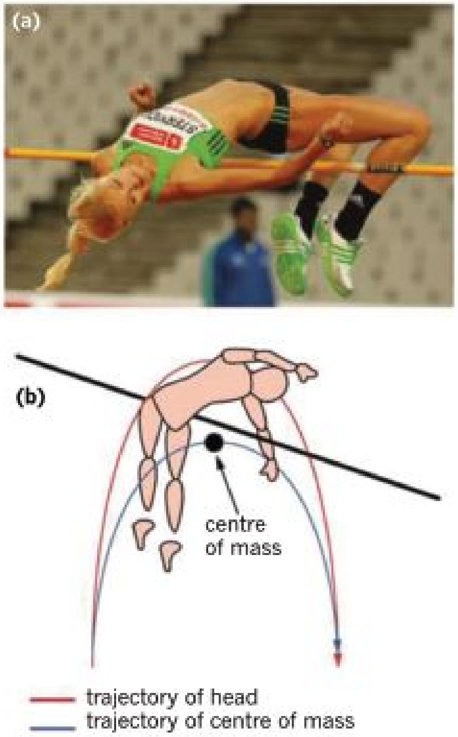


FIGURE 14.1.13 (a) The high jumper clears the bar. Note the position of her legs and head, which lowers her centre of mass and reduces the amount of energy required to complete the jump. (b) This is a diagram of the trajectory of the athlete's head and centre of mass. Note that, while her whole body clears the bar, her centre of mass passes below the bar.

The human body is a complex shape, and so it is often useful to analyse its centre of mass, rather than the whole body, when considering its motion. Remember that, in the case of an object that is not rigid, the mass can redistribute. Figure 14.1.12 shows how the centre of mass moves as the person raises their arms.

The position of the centre of mass is often of interest in sports science. Consider the high jumper pictured in Figure 14.1.13. While her body needs to go over the bar, her limbs move and her back arches in such a way that her centre of mass actually moves under the bar.

Calculation of the centre of mass

In order to calculate the centre of mass of multiple objects, begin by considering two objects which are joined by a rigid but massless wire.

If the positions of the two objects are marked relative to an origin, as indicated in Figure 14.1.14, the position of the centre of mass is defined to be $x_{\text{cm}} = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}$

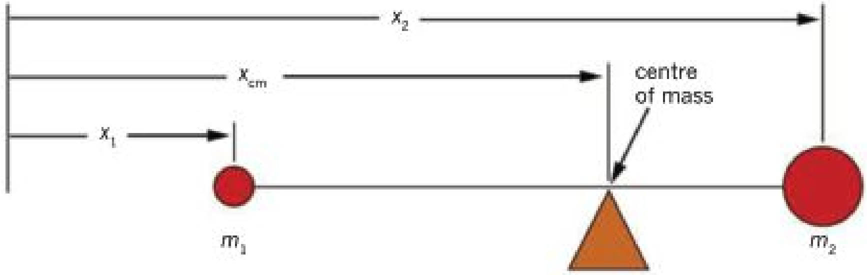


FIGURE 14.1.14 Calculation of the centre of mass. The product of mass and distance is equal for each side of the fulcrum.

In order to confirm that this definition works, suppose, for example, that $m_1 = 0$. In this case, there is only one object, of mass m_2, and the centre of mass must lie at the position of that object, which is x_2. Substituting $m_1 = 0$ into the above equation gives:

$$\begin{aligned} x_{\text{cm}} &= \frac{0 \times x_1 + m_2x_2}{0 + m_2} \\ &= \frac{m_2x_2}{m_2} \\ &= x_2\text{, as expected.} \end{aligned}$$

As a further confirmation, suppose that the two objects have the same mass ($m_1 = m_2$). In this case, the centre of mass would be the average of the positions of the two objects (that is, halfway between the two objects). Replacing both m_1 and m_2 with the same mass, m, gives:

$$\begin{aligned} x_{\text{cm}} &= \frac{mx_1 + mx_2}{m + m} \\ &= \frac{m(x_1 + x_2)}{2m} \\ &= \frac{1}{2}(x_1 + x_2)\text{, as expected.} \end{aligned}$$

It is a simple extension of this principle to calculate the centre of mass for a more complicated arrangement of masses, as shown in Figure 14.1.15.

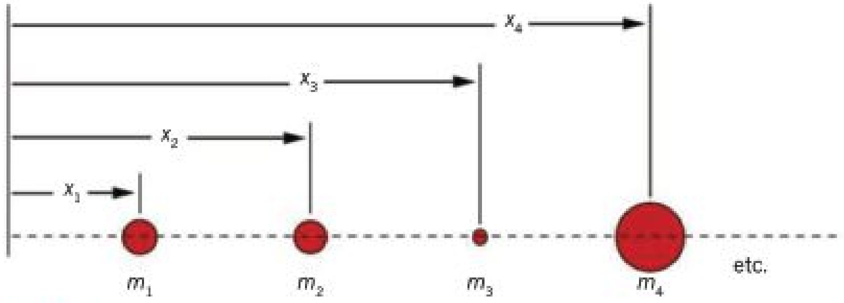


FIGURE 14.1.15 Positions and masses used to find the centre of mass of a collection of masses.

Hence the general formula for the centre of mass for a system of masses placed at positions x_1 to x_n and having masses m_1 to m_n is given by:

$$x_{cm} = \frac{m_1x_1 + m_2x_2 \ldots + m_nx_n}{m_1 + m_2 + \ldots m_n}$$

where m refers to masses 1 to n

x refers to the positions of masses 1 to n relative to the origin

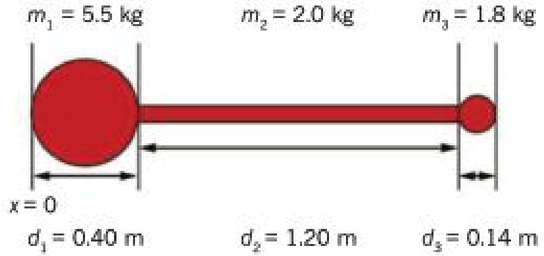
x_{cm} is the position of the centre of mass relative to the origin.

A complex object may be divided into many small pieces, and a computer may be used to calculate the centre of mass to the degree of accuracy required.

Worked example 14.1.1

CALCULATING CENTRE OF MASS

Find the best position to support the object below given the values in the figure. Take the origin on the left as shown and give all responses correct to two significant figures.

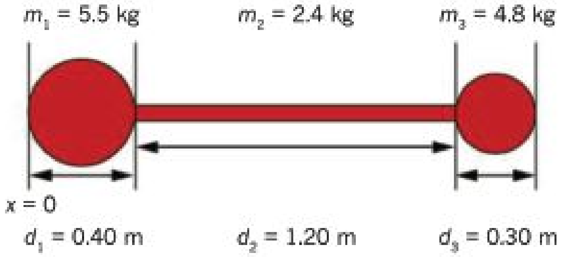


Thinking	Working
Find the centre of mass of each component part, using the symmetry of the body. Remember that while you have been asked to report your answers to two significant figures, you should retain at least one extra significant figure throughout your calculations to avoid rounding errors in your subsequent calculations.	m_1 has centre at x_1 = 0.20 m. m_2 has centre of mass halfway down its length: $x_2 = d_1 + \frac{1}{2}d_2$ $= 0.40\text{ m} + 0.60\text{ m}$ $= 1.0\text{ m}$ m_3 has centre: $x_3 = d_1 + d_2 + \frac{1}{2}d_3$ $= 0.40\text{ m} + 1.20\text{ m} + 0.07\text{ m}$ $= 1.67\text{ m}$
Now substitute into the centre of mass equation: $x_{cm} = \frac{m_1x_1 + m_2x_2 \ldots + m_nx_n}{m_1 + m_2 + \ldots m_n}$	$x_{cm} = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3}$ $= \frac{5.5(0.20) + 2.0(1.0) + 1.8(1.67)}{5.5 + 2.0 + 1.8}$ $= 0.66\text{ m}$
The object is best supported at the centre of mass.	The object is best supported at the centre of mass, 0.66 m from the origin.
Always check if your answer is reasonable.	Because there is more mass concentrated on the left-hand side, one would expect x_{cm} = 0.66 m to be to the left of the middle (x = 0.87 m).

Worked example: Try yourself 14.1.1

CALCULATING CENTRE OF MASS

Find the best position to support the object below given the values in the figure. Take the origin on the left as shown and give all responses correct to two significant figures.



The centre of mass approach can be useful in studying the movement of the human body. Centres of mass of body parts can be found separately, and then the relative movement of the parts can be considered. For example, the torso and the head could be considered as two separate centres of mass when studying whiplash in collision simulations.

The centre of mass and stability

For a body to be stable, the total torque on it must be zero. This condition implies that the centre of mass must lie above the base of support. When a person attempts to touch their toes, their centre of mass moves forwards. If they cannot lean backwards to compensate, their centre of mass falls in front of their feet. This produces a torque that tends to make them fall forwards. A similar example is shown in Figure 14.1.16. In general, the body is stable as long as the centre of mass lies above the base of support. As the mass distribution of our bodies changes when we are in motion, we instinctively move in an attempt to maintain our balance. Centre of mass and stability were discussed in more detail in Section 11.3.

FIGURE 14.1.16 This wake-boarder is no longer able to keep their centre of mass above their feet, and so they overbalance.

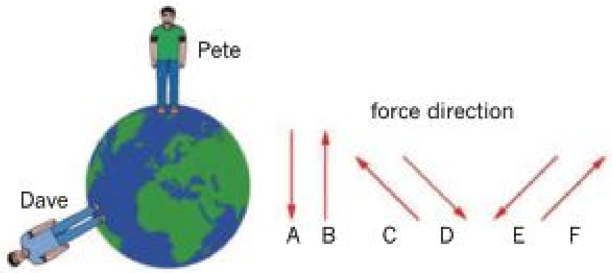
14.1 Review

SUMMARY

- Contact forces are directly applied to a body.
- Gravity is a non-contact force that, together with the normal reaction force, causes compression of the bones in the weight-bearing structures of the body.
- A material that is being stretched by forces acting on it is under tension. Stretching forces are called tensile forces.
- A material that is being squeezed as a result of forces acting on it is under compression. Squeezing forces are known as compressive forces.
- Shear is caused by a pair of sideways forces acting in opposite directions. This causes the shape of the material to alter.
- As a material bends, compressive, tensile and shear forces are all acting together.
- The net external force acting on a body can be considered to act on its centre of mass.
- The centre of mass is the point at which the mass of a body may be considered to be concentrated.
- The centre of mass of a system of masses m_i located at positions x_i may be calculated by:
  $$x_{cm} = \frac{m_1x_1 + m_2x_2 \ldots + m_nx_n}{m_1 + m_2 + \ldots m_n}$$
- The centre of mass of the human body moves as the body position changes.
- For stability, the centre of mass must lie above the base of support.

KEY QUESTIONS

1 The figure below shows Pete and Dave standing on the surface of the Earth.



Choose the force directions A–F that most closely match the forces described:

a the normal reaction force on Pete

b the force due to gravity on Earth by Pete

c the force due to gravity on Pete by Earth

d the normal reaction force on Dave.

2 Gravity on the surface of Mars is just over one-third of that on Earth. Choose the statement(s) that are true about the forces on an astronaut on Mars compared to Earth.

A The normal reaction force on the astronaut would be greater than on Earth.

B The normal force on the astronaut would be less than on Earth.

C The astronaut would experience greater compressive forces on load-bearing bones.

D The astronaut would experience smaller compressive forces on load-bearing bones.

3 Select the person who is not in translational equilibrium.

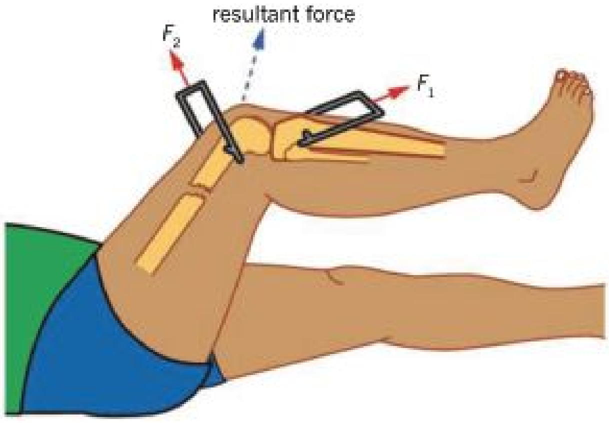
A a motorist going around a curve in the road in her car at constant speed

B a man moving upwards in an elevator at constant speed

C a child floating in a swimming pool

D a parachutist floating down towards Earth at terminal velocity

4 The bones below are secured with pins and subjected to the forces shown. What type of force is exerted on the fractured femur (thigh bone)? Choose the best answer.



A compression

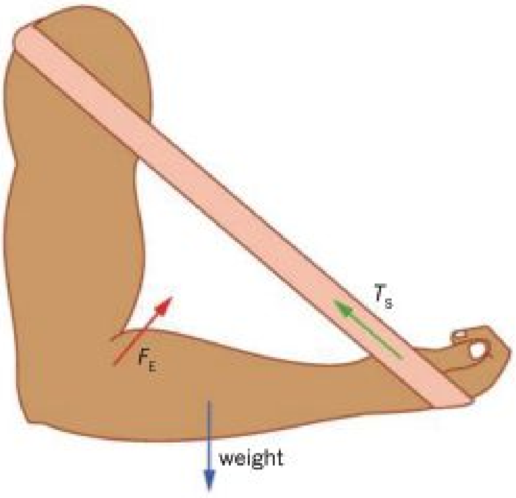
B tension

C shear

D bending

14.1 Review *continued*

5 Consider the sling designed to temporarily support a fracture in the forearm. The force F_E is the effective force on the forearm at the elbow joint from the bones and tendons, while T_S is the force exerted by the sling. What type of force is exerted on the fracture? Choose the most suitable answer.

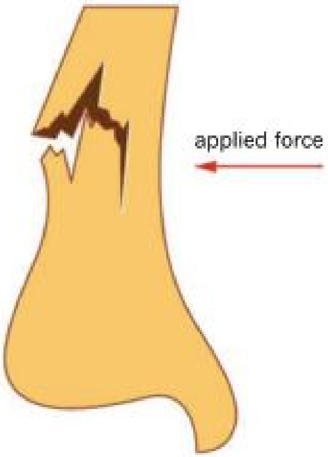


A shear
B tension
C compression
D twisting

6 Where, approximately, is the centre of mass for the human body when standing upright?

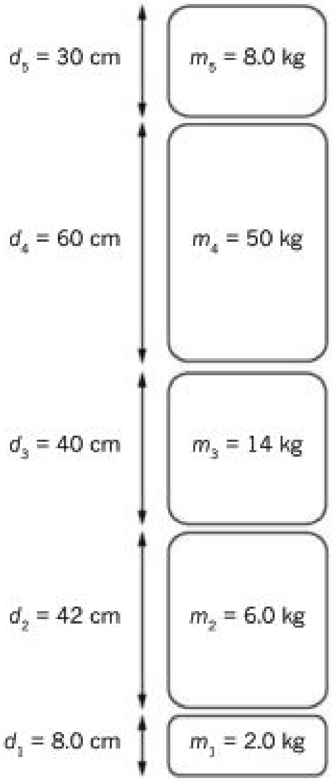
7 In many sports it is important to be able to maintain stability on your feet. What actions could you employ to increase your stability?

8 Explain why the greenstick fracture below has started at the side of the bone opposite the applied force.

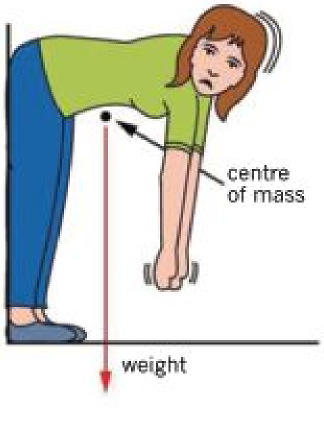


9 A new material being considered for bone replacement is stronger under tension than compression. Describe how an artificial femur made of this material would be likely to fracture if the femur was struck with a heavy object from the side.

10 Consider a person with a height of 180 cm and a mass of 80 kg. Assume that you can simplify the person so they are made up of five symmetrical parts and each part has a uniform mass distribution. The parts are: feet, with height 8.0 cm and mass 2.0 kg; lower legs, with height 42 cm and combined mass 6.0 kg; upper legs, with height 40 cm and combined mass 14 kg; torso, with height 60 cm and mass 50 kg; and head, with height 30 cm and mass 8.0 kg. The diagram below is a schematic representation of this person. Use this information to calculate the height of the person's centre of mass when they are standing upright.



11 Explain in your own words why the centre of mass of the girl shown lies outside her body, and why she is unstable as she is currently standing.



14.2 Forces cause rotation

Football players, such as those in Figure 14.2.1, kick footballs. In training sessions, athletes may lift hand weights repeatedly and perform sit-ups and chin-ups. When body parts move in these ways, they rotate about joints. Rotation about the joints is an important effect of the forces on the body and is the basis for most mobility (movement). The physics of levers is useful to model the behaviour of joints when forces are applied by loads, muscles and tendons.

FIGURE 14.2.1 When playing sports, joint tissues and limb bones are subjected to a variety of forces which can cause rotation.

TORQUE

Forces caused by muscles produce a turning effect or torque. For an object to turn, there needs to be a pivot point around which the object will rotate. In the body, these points are the joints. The muscles apply a force to the bones via tendons in such a way as to cause the bones to rotate about the joint. In this way, the bones or limbs being rotated can be referred to as the **lever**. Note that the length of the lever refers to the distance between the pivot and the point of application of the force and is given by the symbol r. These points are labelled on the diagram in Figure 14.2.2.

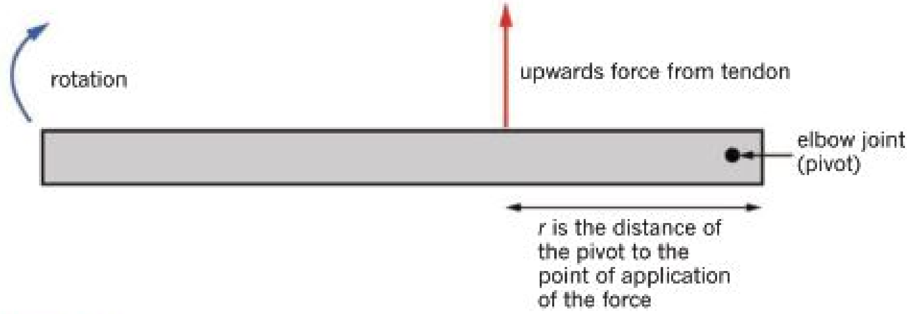


FIGURE 14.2.2 In this schematic diagram of the human elbow, the force is applied upwards on the forearm and the arm rotates about the elbow joint. The distance from the elbow joint to the point of application on the forearm is shown. This is the lever arm.

The force applied cannot go through the joint, or it would not cause rotation (see Figure 14.2.3).

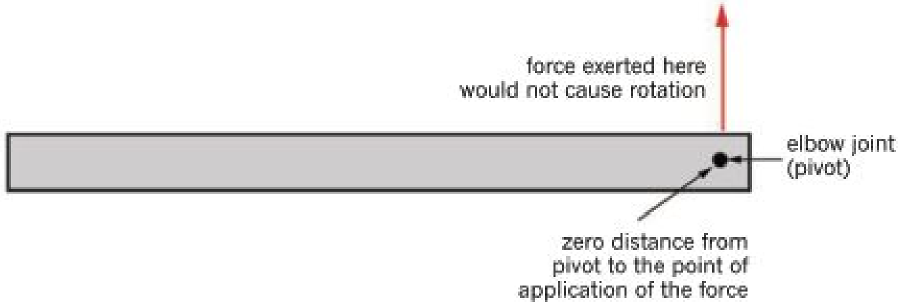


FIGURE 14.2.3 When the force is applied through the pivot point, there can be no rotation. A tendon attached to the elbow itself would cause the forearm to lift but not to rotate.

When the force acts perpendicularly to the lever arm, the torque is simply calculated as the product $\tau = rF$. Torque is measured in newton metres (N m).

An object will be in rotational equilibrium when the torques applied to it are equal but in opposite directions. In this case, the object will not rotate. If there is a resultant torque, however, the object will rotate.

The torque depends on the size of the force causing rotation, the angle at which the force is applied, and the length of the lever arm. In Figure 14.2.4 on page 500, the force is applied at an angle, and not perpendicular to the forearm as in Figure 14.2.2. When the force is applied at an angle, it is only the perpendicular component of the force, $F_{\perp} = F \sin \theta$, that has any effect. The parallel component acts on a line along the lever, and through the pivot. It does not cause rotation.

Therefore the torque (τ) acting on a body is given by the product of the component of the lever arm, r, and the perpendicular component of the applied force.

$\tau = rF_{\perp}$

where τ is the torque (in Nm)

r is the distance between the pivot and the point of application of the force (in m)

$F_{\perp}$ is the perpendicular component of the force applied (in N).

When the applied force makes an angle of θ with the lever arm, then:

$\tau = rF \sin \theta$

This rule is explained in Figure 14.2.4.

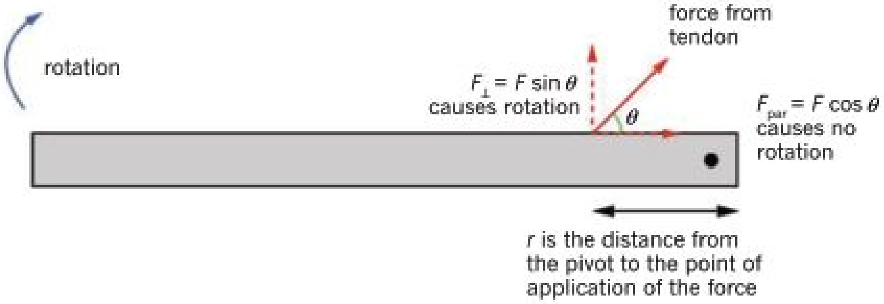


FIGURE 14.2.4 The force is applied at an angle to the forearm. Only the perpendicular component ($F \sin \theta$) generates rotation. The component of force acting through the pivot does not cause rotation.

Torque in the body

The muscle group commonly called the hamstring attaches to the tibia and fibula of the lower leg. When these muscles contract, they pull up on the rear side of the tibia and fibula. This creates an anticlockwise torque when viewed from the left side, as depicted in Figure 14.2.5. The contraction of these muscles causes the knee to flex (bend). The weight attached to the foot exerts a force that tends to rotate the lower leg clockwise, as shown in Figure 14.2.5(a). In Figure 14.2.5(a), the weight is exerting maximum torque. In Figure 14.2.5(b), the lower leg has already rotated. In this case, the weight attached to the foot is still directed vertically downwards, but the component of the force acting perpendicular to the leg ($F \sin \theta$) is less.

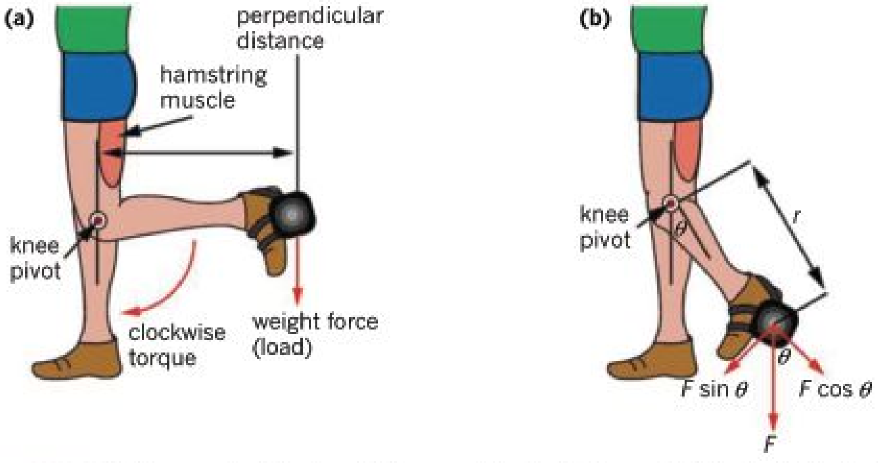
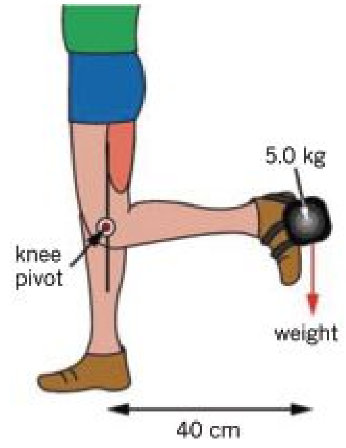


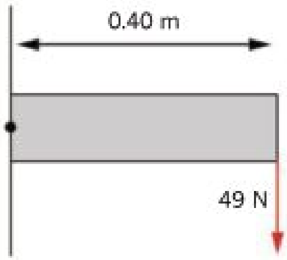
FIGURE 14.2.5 (a) Torque about the knee joint is caused by the training weight attached to the foot. (b) The component of force acting perpendicular to the leg is less when the leg has rotated.

Worked example 14.2.1

CALCULATING TORQUE

The weight attached to the ankle has mass 5.0 kg, and it is placed 40 cm from the knee joint. The lower leg is at a right angle to the vertical. Calculate the torque generated by the weight.



Thinking	**Working**
Convert to SI units.	$r = 40\text{ cm} = 0.40\text{ m}$
Calculate the weight using $g = 9.8\text{ N kg}^{-1}$.	$F = mg$ $= 5.0\text{ kg} \times 9.8\text{ N kg}^{-1}$ $= 49\text{ N}$
Draw a schematic diagram of the forces.	
Check if the forces acting are perpendicular to the body being rotated. If not you need to find the perpendicular component.	In this case F is perpendicular to the leg.
Identify the distance from the pivot point to the point of application of the force.	The force is applied 0.40 m from the pivot.
Calculate the torque.	$\tau = rF \sin 90°$ $= 0.40\text{ m} \times 49\text{ N} \times \sin 90°$ $= 20\text{ N m}$

Worked example: Try yourself 14.2.1

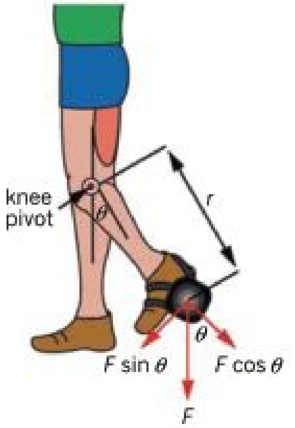
CALCULATING TORQUE

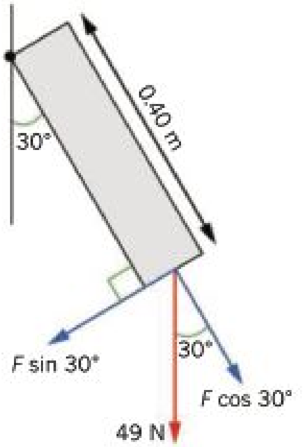
The person in Worked example 14.2.1 gets tired and decides to reduce the mass to 4.0 kg. It is still placed 40 cm below the knee, and the leg is at a right angle to the vertical. Calculate the torque generated by the weight. (For all your calculations, report your answers correct to two significant figures.)

Worked example 14.2.2

CALCULATING TORQUE FROM A SPECIFIC ANGLE

The weight attached to the ankle has mass 5.0 kg, and it is placed 40 cm from the knee joint. The lower leg is at 30° to the vertical. The angle is marked as θ in the diagram. Calculate the torque generated by the weight.



Thinking	Working
Convert to SI units and calculate the weight.	$r = 40 \text{ cm} = 0.40 \text{ m}$ $F = mg$ $= 5.0 \text{ kg} \times 9.8 \text{ N kg}^{-1}$ $= 49 \text{ N}$
Draw a schematic diagram of the forces.	
Check if the forces acting are perpendicular to the body being rotated. If not you need to find the perpendicular component.	$F \sin 30°$ is the component of the force perpendicular to the lever arm.
Find the distance from the pivot point to the point of application of the force.	The force is applied 0.40 m from the pivot.
Calculate torque using $\tau = rF \sin \theta$.	$\tau = rF \sin \theta$ $= 0.40 \text{ m} \times 49 \text{ N} \times \sin 30°$ $= 9.8 \text{ N m}$

Worked example: Try yourself 14.2.2

CALCULATING TORQUE FROM A SPECIFIC ANGLE

The weight attached to the ankle has mass 4.0 kg, and it is placed 40 cm from the knee joint. The lower leg is at 45° to the vertical. Calculate the torque generated by the weight.

JOINTS AS SIMPLE LEVERS

Bones pivot about joints when acted on by muscles and tendons. They are classified as class 1, class 2 or class 3 levers, depending on where the pivot point is relative to the **load** and the **effort** (applied force). Levers can function as simple machines that allow the relatively small forces exerted by muscles to be multiplied. Alternatively, levers can multiply speed, allowing small movements of a muscle to create larger movements for a limb, for instance. The forces applied to the lever provide the torque that rotates the bones at the joints in the human body.

Class 1 lever

The classic class 1 lever is like a see-saw, where the effort and load are on opposite sides of the pivot point. If the distance between the effort force and the pivot point is greater than the distance between the load and the pivot point, the effort force required will be less than the load. In this case, the class 1 lever is acting as a **force multiplier**. This is shown in Figure 14.2.6. On the other hand, if the distance between the effort and the pivot point is less than the distance between the load and the pivot point, then the effort force required will be greater than the load. In this case the point of application of the load will travel further, and therefore faster, than the point of application of the effort, and so the lever is acting as a **speed multiplier**.

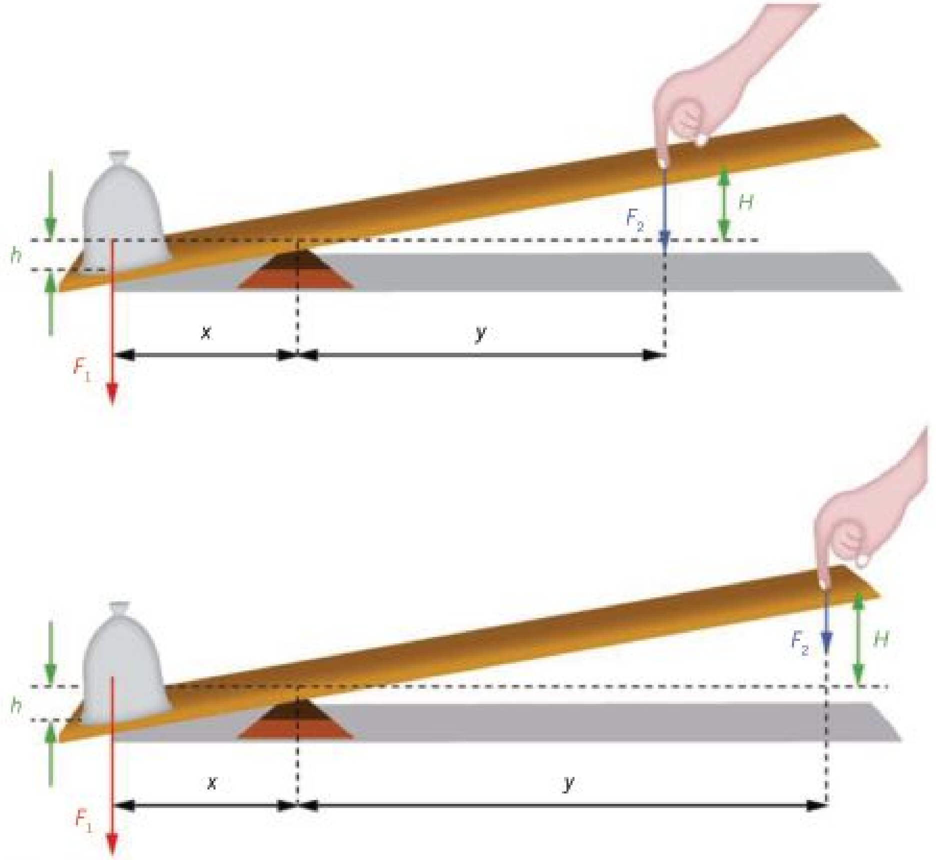


FIGURE 14.2.6 A class 1 lever. In this diagram F_1 represents a load force and F_2 is the effort force. By increasing y, the distance between the effort force and the pivot point, we reduce the effort force required to lift the load.

In the human body, the point where the skull meets the spine is a pivot point, as shown in Figure 14.2.7. The mass of the head acts in front of the pivot, but the muscles at the back of the neck provide a balancing force to keep the head up. When a person relaxes these muscles and slowly contracts them again, the head nods. For the head to remain up, the muscles have to provide a constant force to balance the weight of the head. When the head is stationary, the total torque about the pivot point is zero.

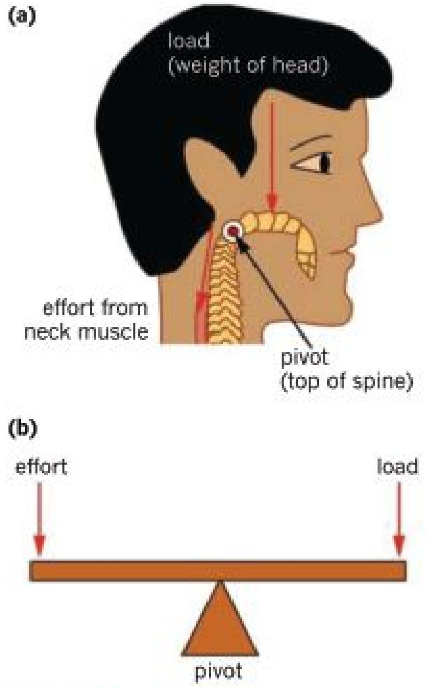


FIGURE 14.2.7 (a) The skull is attached to the first vertebra of the neck. A muscle in the neck is constantly contracted to supply the effort force that counters the load of the head's weight on the other side of the pivot. (b) The class 1 lever arrangement that has the pivot between the effort and the load.

Class 2 lever

When a person stands on their toes (pivot), their weight (load) acts effectively in a line through the arch of the foot. The effort is supplied by the contraction of the calf muscles, which are attached to the heel by the Achilles tendon. The load and the effort act on the same side of the pivot point, as shown in Figure 14.2.8. This is a class 2 lever, which always acts as a force multiplier.

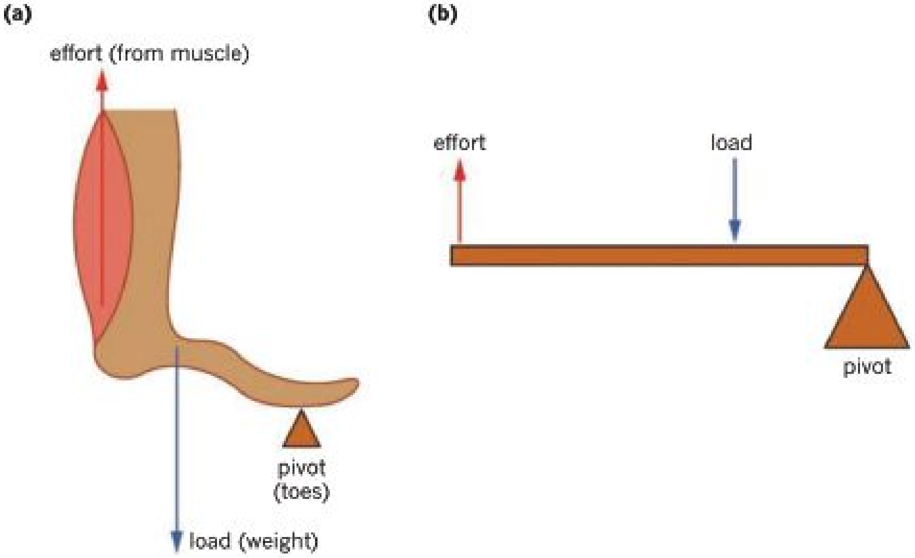


FIGURE 14.2.8 (a) The toes form the pivot point. The Achilles tendon supplies an effort force that lifts the load, which is the weight of the person transferred by the leg bones. (b) Class 2 lever with effort and load on the same side of the pivot, with the load closer to the pivot.

There is a **mechanical advantage** generated by the fact that the tendon is attached further from the toes than the distance between the weight and the toes. For this reason the muscle is able to lift the body. Note, too, how people tend to lean forwards when standing on their toes. This further reduces the load-to-toes distance and reduces the force required from the calf muscles.

Class 3 lever

In a class 3 lever, the effort is supplied between the pivot and load. The elbow joint shown in Figure 14.2.9 is the pivot point, with the effort supplied by the biceps muscle. The load is the weight of the arm and anything that is being lifted in the hand. While this type of lever does not generate a mechanical advantage, it does allow a small movement of the muscle to produce a large movement of the forearm. Class 3 levers always function as speed multipliers.

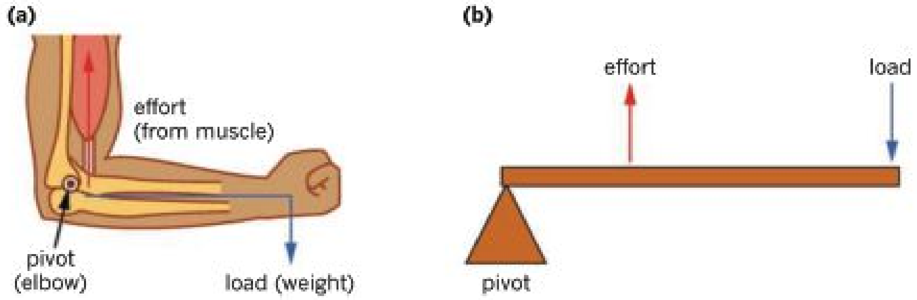


FIGURE 14.2.9 (a) Class 3 lever: the load is the weight of the forearm itself and anything held in the hand. The effort is provided by the biceps and the pivot is the elbow joint. (b) Class 3 lever with the effort between the pivot and the load.

14.2 Review

SUMMARY

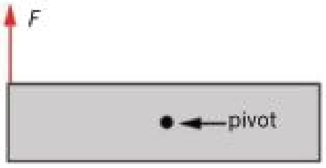
- Forces can cause rotation at joints, with the torque being given by $\tau = rF$.
- When the force is applied at an angle θ to the lever arm, only the component perpendicular to the arm causes torque, so $\tau = rF \sin \theta$. Joints in the human body may be modelled as levers.
- Levers have a pivot point, which is the joint itself.
- The effort in the lever is provided by a muscle attached to a tendon.
- The load is the weight of the body part, and any extra weight attached.
- The lever may act as a force multiplier, such as at the ankle, where the force of the calf muscle is multiplied to lift the body when standing on the toes.
- The lever may act as a speed multiplier, for example, when a small movement of the biceps muscle multiplies to flex the forearm.

KEY QUESTIONS

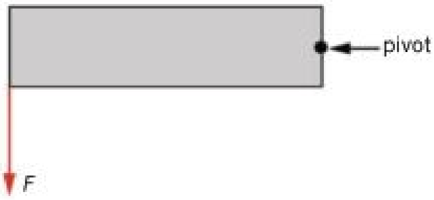
1 Which of the following would experience the greatest torque?

A

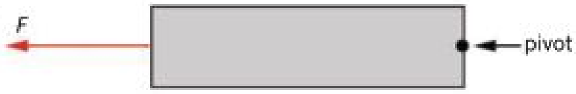
B



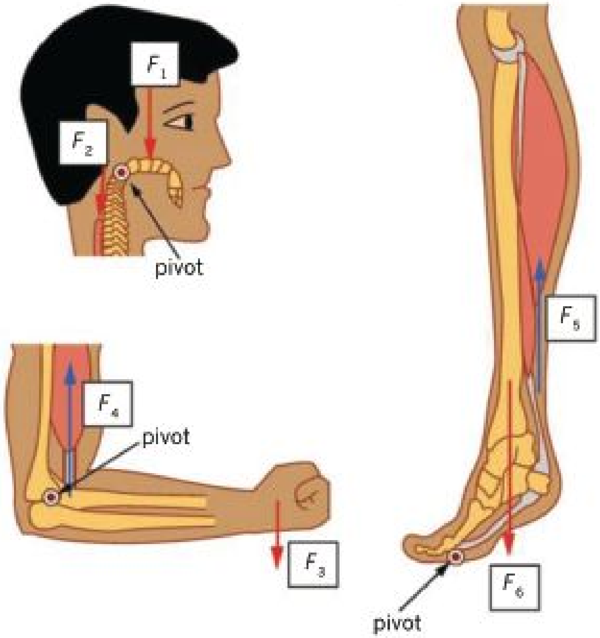
C



D



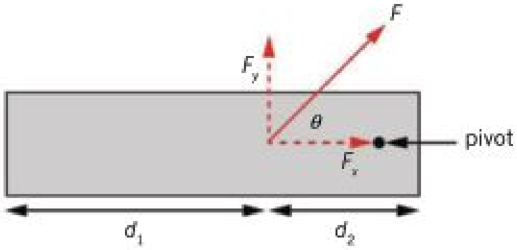
The following diagrams are relevant to questions 2 and 3.



2 For each of the diagrams, choose which force vector represents the effort and which represents the load.

3 Classify the skull–neck joint, the elbow joint, and the toe joint as either class 1, class 2 or class 3 levers, and state whether they act as force multipliers or speed multipliers.

4 Write down an expression for the torque around the pivot point in the diagram.



14.2 Review *continued*

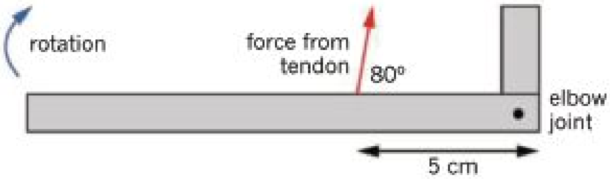
5 Choose the correct statement(s).
The magnitude of a torque caused by force *F* can be increased by:

A applying force *F* closer to the pivot
B applying force *F* at 90° to the lever arm
C applying the force further from the pivot
D applying the force along the lever arm

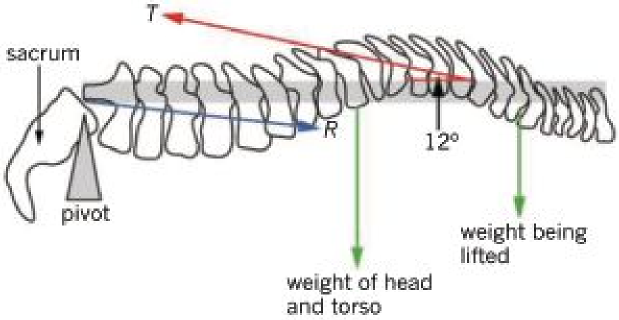
6 A person holds a 12 kg training weight in his right hand. His forearm is held in a horizontal position with no rotation at the elbow joint. The weight is 0.38 m from the elbow joint and the tendon joins the forearm 5.0 cm from the joint, at an angle of 90°.

a Calculate the torque generated by the weight.
b How much muscle force is required to keep the weight from rotating the elbow joint? (That is, over and above the force that would be required to keep the hand and arm in the same position.) Answer to two significant figures.

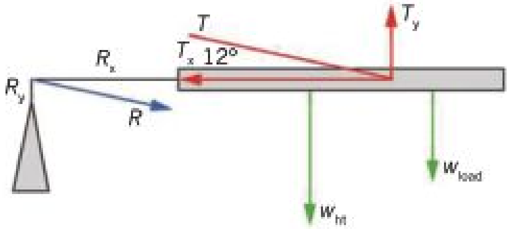
7 Researchers measured the torque produced at the elbow joint using a specially adapted brace that could measure the torque produced by the muscles as a patient attempted to flex their elbow. Patients were required to exert the maximum force of which they were capable, trying to reduce the angle at the elbow. A patient with muscular dystrophy achieved a torque of 1.34 N m. The tendon attaches to the forearm 5.0 cm from the elbow joint, and in this configuration it exerts a force at 80° to the forearm, as shown below. Calculate the force produced by the biceps muscle correct to two significant figures.



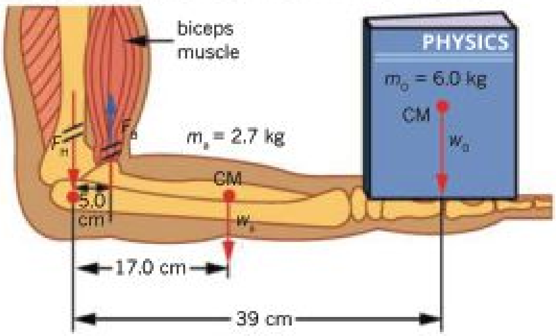
8 The diagram below shows the forces on the spine when a heavy object is picked up by bending at the hip. The sacrum is the pivot point. The force marked *R* is the reaction force of the sacrum on the 5th lumbar vertebra, and is considered to act horizontally. The force *T* is the tension provided by the back muscles, which are attached further up the spine. This force acts at 12° to the spine. By considering the torque about the sacrum, explain why this lifting technique often leads to injury. Hint: consider the size of the force *T* and the type of stress placed on the back.



The components of these forces can be represented as follows:



9 A person is supporting a 6.0 kg book in his hand with his elbow flexed at 90°, as shown in the diagram below. The force F_H exerted by the humerus at the pivot point is shown. The lower arm has a mass of 2.7 kg and the weight can be considered to act 17 cm from the elbow. The 6.0 kg book is 39 cm from the elbow. The biceps muscle attaches 5.0 cm from the elbow, and exerts a force F_B.
Calculate the force that has to be exerted by the biceps muscle to keep the arm in this position.



14.3 Tissue under load: Stress and strain

The therapeutic and cosmetic use of artificial materials for reconstruction of the human body is on the rise, and many young people today will have joints and possibly other body parts replaced as they age. Figure 14.3.1 shows an artificial knee replacement that is a common procedure among sports people and the ageing population. It is vital that these replacements have physical properties that make them as good as, or superior to, the biological tissues they replace.

When forces act on a material it is said to be under **stress**. These stresses can be compressive, tensile, shear or a combination of these. Many tissues in the human body, as well as common building materials, respond differently to these different types of forces. Just how much stress the material can withstand without failing gives an indication of its strength. Many materials deform when they are compressed or stretched. This is described by the strain that quantifies the change in dimensions relative to the undistorted material. Both stress and strain are important parameters in understanding the behaviour of the materials in the human body and in seeking materials for **prostheses**.


FIGURE 14.3.1 An illustration of an artificial knee replacement.

CALCULATING STRESS

When subject to the same force, a thin elastic band stretches more than a thick one, as shown in Figure 14.3.2. This is because of the physical geometry of the elastic band, rather than the nature of the rubber. In order to study the response of rubber to an external force, it is useful to define the stress as the force per unit area of cross section. The stress measured in this way is independent of the dimensions of the sample being tested. The symbol for stress is σ (the Greek letter sigma).

The stress (σ) applied to a material is the load force per unit area of cross-section. The load force may place the material under compression or tension:

$$\sigma = \frac{\text{load force}}{\text{cross sectional area}} = \frac{F}{A}$$

where F is the force measured in newtons (N)

A is the area measured in square metres (m^2)

σ is the stress measured in newtons per square metre (N m^{-2}).

Note: 1 newton per square metre is equal to 1 pascal (Pa).

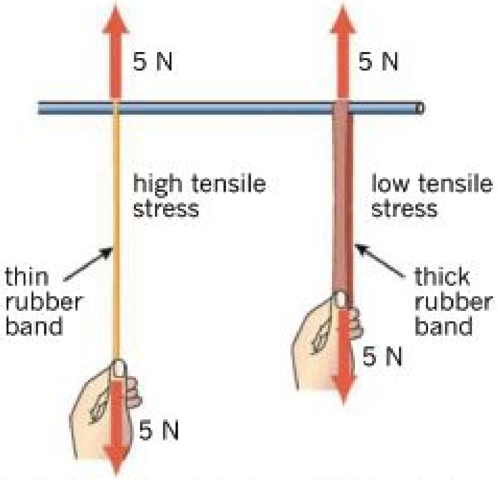


FIGURE 14.3.2 The tension in these rubber bands is equal. A 5 N stretching force has been applied to both rubber bands. The stress experienced by each rubber band, however, is not equal. The thin rubber band is under more stress than the thick rubber band because of its smaller cross-sectional area.

In the body, the force applied can put a material under tension, which causes stretching, or forces can be compressive, which causes squashing of the material. Shear stresses are more complicated to calculate, but most bone fractures are the result of shear stress, and so it is important to have data on the performance of biological materials under shearing forces. The same formula for calculating stress applies to all types of stress.

Worked example 14.3.1

CALCULATING STRESS

Calculate the stress in an elastic band of cross-sectional area 1 mm^2 which is subject to a 5 N force.

Thinking	Working
Convert the area to SI units.	1 mm = 0.001 m, hence $1\ mm^2 = (0.001)(0.001)$ $= 1 \times 10^{-6}\ m^2$
Use the formula $\sigma = \frac{F}{A}$	$\sigma = \frac{5\ N}{1 \times 10^{-6}\ m^2}$ $= 5 \times 10^{6}\ N\ m^{-2}$

Worked example: Try yourself 14.3.1

CALCULATING STRESS

Calculate the stress in a thicker elastic band of cross-sectional area 5 mm^2 which is subject to a 5 N force.

Note that the stress is significantly less in the thicker rubber band.

STRENGTH

The tibia is one of the strongest load-bearing parts of the body. It is possible to calculate just how much load it can take and at what point it would fracture. The **strength** is the maximum stress that the material can withstand before it fails. A bone would have different strength under compressive, tensile and shearing forces.

Since strength is simply the maximum stress, strength is also measured in pascals (Pa) or newtons per square metre ($N\ m^{-2}$).

The strengths of some important engineering and biological materials are shown in Table 14.3.1.

Material	Compressive strength (MPa)	Tensile strength (MPa)
cast iron	550	170
steel	500	820
aluminium	200	200
timber (pine)		
• along the grain	35	40
• across the grain	10	
nylon fibre	0	500
concrete	20	2
carbon fibre	0	6000
carbon nanotube	0	14000
bone	170	130
concrete	20	2
tendon		100
tooth enamel	380	35

TABLE 14.3.1 Typical compressive and tensile strengths for a variety of materials.

STRAIN AND DEFORMATION

When a material is under stress, it stretches or squashes. Microscopically, the atom-to-atom distance will change when the load is applied. This changes the macroscopic dimensions. We call this distortion caused by stress the **strain**.

To be more precise, it is the fractional change in the length of a material under stress. Strain is represented by the Greek letter epsilon (ε) and is always expressed as a positive value, regardless of whether the stress is tensile or compressive. Since strain is the ratio of two lengths, it has no unit. It can be represented as a decimal number (e.g. 0.02) or a percentage (e.g. 2%).

The strain (ε) is the amount of distortion (i.e. extension or compression) per unit length of the material:

$$\varepsilon = \frac{\text{change in length}}{\text{original length}} = \frac{\Delta l}{l}$$

Strain is dependent on the type of material. For example, if a person were hanging off a length l of bungee rope with cross-sectional area A, it would stretch significantly, as shown in Figure 14.3.3. However, if the same person were to hang off a length l of steel cable with the same cross-sectional area A, the cable would hardly stretch at all. Ignoring the masses of the cable and rope, the tension force and the tensile stress in both the cable and the rope are the same—given by the weight of the person and the cross-sectional area of the material. The *effect* of the tensile force on each material is very different. In an application such as bungee jumping, it is critical to predict Δl correctly!

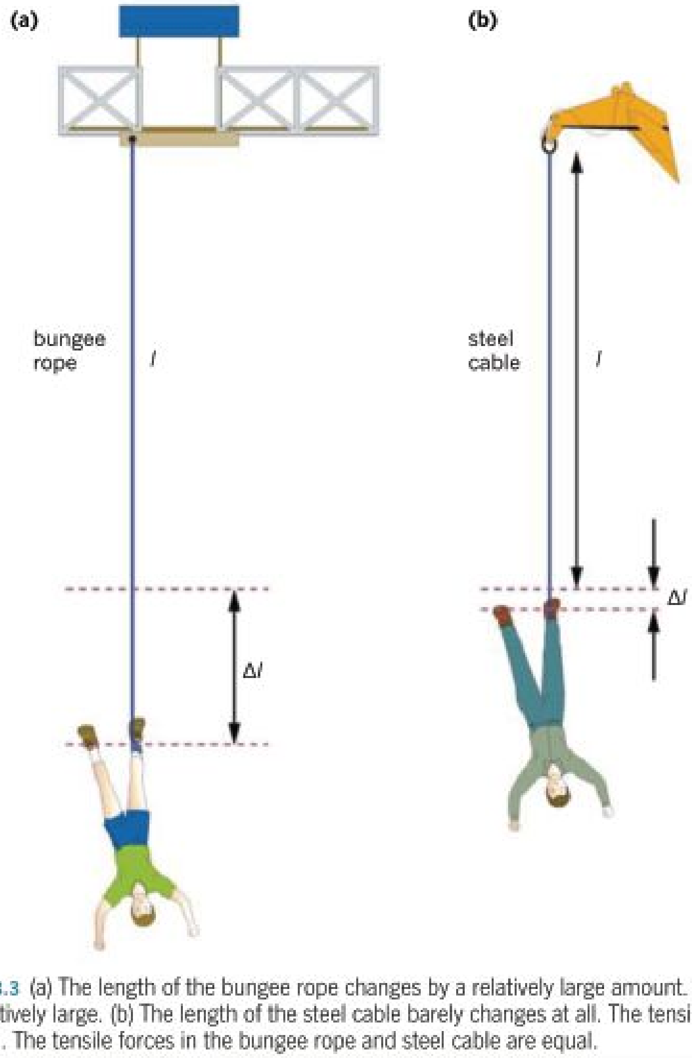


FIGURE 14.3.3 (a) The length of the bungee rope changes by a relatively large amount. The tensile strain is relatively large. (b) The length of the steel cable barely changes at all. The tensile strain here is very small. The tensile forces in the bungee rope and steel cable are equal.

Worked example 14.3.2

CALCULATING STRAIN

While preparing for training, an athlete stretches her hamstrings. Her hamstrings are usually 25 cm in length, but extend to a length of 26 cm while being stretched. Calculate the strain.

Thinking	Working
Calculate the change in length, Δl.	$\Delta l = 26\text{ cm} - 25\text{ cm}$ $= 1.0\text{ cm}$
Use the formula $\varepsilon = \frac{\Delta l}{l}$ Note that you do not need to convert to SI units, as long as both length measurements use the same unit.	$\varepsilon = \frac{1.0}{25}$ $= 0.040$

Worked example: Try yourself 14.3.2

CALCULATING STRAIN

While training, the athlete contracts her hamstrings. Her hamstrings are usually 25 cm in length, but shorten to a length of 23 cm while being contracted. Calculate the strain.

Worked example 14.3.3

CALCULATING STRAINED LENGTH

While running, an athlete's calf muscle undergoes a tensile strain of 0.030. The original length of the muscle is 30 cm. Calculate the length of the muscle while it is being strained. Give your answer to two significant figures.

Thinking	Working
Rearrange the following formula to make Δl the subject: $\varepsilon = \frac{\Delta l}{l}$	$\Delta l = \varepsilon l$
Substitute the given values to calculate Δl. Note that you do not need to convert to SI units. The answer will be in the same units as the length.	$\Delta l = 0.030 \times 30$ $= 0.90\text{ cm}$
Calculate the strained length, remembering that the calf is undergoing tensile strain, which means the new length will be longer than its original length.	Strained length = 30 + 0.90 = 30.90 cm = 31 cm (to two significant figures)

Worked example: Try yourself 14.3.3

CALCULATING STRAINED LENGTH

While running, the same athlete's calf muscle undergoes a compressive strain of 0.040. The original length of the muscle is 30 cm. Calculate the length of the muscle while it is being strained. Give your answer to two significant figures.

When considering the human body, the deformation in response to stress is very important. Muscles contract in order to place tendons under tension, which makes our bones move. Tendons therefore need to be able to stretch slightly and then return to their normal rest length. Bones need to be able to sustain various types and magnitudes of stress when subject to the forces generated by movement.

14.3 Review

SUMMARY

- Stress is the force per unit area in a material.
- In general, stress can be tensile, compressive or shear stress, and the stress properties of a specific material may be different for each type of stress.
- Stress, $\sigma = \frac{\text{load force}}{\text{cross sectional area}} = \frac{F}{A}$
- Strength is the maximum stress that a material can withstand before it fails.
- When a force acts on a material, the material deforms.
- The ratio of the change in length to the original length is called the strain, ε.
- Strain, $\varepsilon = \frac{\text{change in length}}{\text{original length}} = \frac{\Delta l}{l}$, is dimensionless, and may be expressed as a percentage.

KEY QUESTIONS

1 Consider two bones: bone 1 has a cross-sectional area of 225 mm^2, and bone 2 has a cross-sectional area of 400 mm^2. Assume the bone material is identical in both bones. Which of the following statements is correct?

A Bone 2 has the greater strength because it can withstand the greater force.
B Bone 2 has the greater strength because it can withstand the greater stress without failing.
C Bone 1 and bone 2 have the same strength.
D There is insufficient information to tell which bone has the greater strength.

2 An adult bone of area A is subject to a force F, causing a stress σ. If an infant bone with area 0.5 A is subjected to the same force, what is the stress in this bone?

3 **a** When a person stands on their toes, the tension in the Achilles tendon is 2.2×10^3 N. Find the stress in the tendon if the cross-sectional area of the tendon is 57 mm^2.

b A laboratory test shows that tendons fail at a stress of 1.1×10^8 N m^{-2}. Calculate the minimum force in the Achilles tendon described above that would cause this failure. Answer correct to two significant figures.

4 A typical Achilles tendon has an average length of 180 mm and can be expected to fail at 12% extension. Calculate the length of the tendon just before it fails. Give your answer correct to two significant figures.

The following information relates to questions 5 and 6. When an 85 kg man stands on one leg, putting his whole body weight on the leg, his femur shortens by 0.15 mm. The femur is originally 450 mm long, and has diameter 35 mm.

5 Calculate the strain in the femur. Give your answer correct to two significant figures.

6 Calculate the stress in the femur. Give your answer correct to two significant figures.

7 Select the correct statement.

A Stress is dimensionless and strain is measured in Pa or N m.
B Stress is measured in N m and strain is dimensionless.
C Stress is measured in Pa or N m^{-2} and strain is measured in N.
D Stress is measured in Pa or N m^{-2} and strain is dimensionless.

8 A 3.1 kg mass is hung from the end of a tendon sample in a laboratory. The sample has a cross-sectional area of 9.0 mm^2. If the tendon stretches by 4% of its rest length, calculate the stress and the strain in the sample. Give your answer correct to two significant figures and use $g = 9.8$ m s^{-2}.

9 Two samples are made from the same materials. They have the same cross-sectional area, but they each have a different length. Choose the correct statement(s).

A If both samples are subjected to the same force, they will experience the same stress.
B If both samples are subjected to the same force they will experience the same strain.
C If both samples are subjected to the same force, they will experience different stress and strain.
D Both samples will deform by the same amount if subjected to the same force.
E Both samples will deform by the same relative amount if subjected to the same force.

10 A muscle exerts a force of 1.8×10^3 N on a tendon of cross-sectional area 25 mm^2, and stretches it by 5%. Calculate the stress and the strain in the tendon.

14.4 Properties of human tissues

FIGURE 14.4.1 The rupture of the Achilles tendon is one of the most common injuries in sports.

Materials under stress tend to deform. In the human body, tendons stretch under daily operation, but the change is not permanent. Excessive stress causes permanent changes at the tissue level, and this often results in injury (see Figure 14.4.1). In this case, some energy is absorbed by the tissue which is not recovered when the stress is removed. The study of the stress–strain relationship is useful in characterising the properties of various materials, and in predicting the loads at which damage or failure will occur.

STRESS-STRAIN GRAPHS

For many materials, the strain increases in proportion to the applied stress, so long as the stress is not too large.

The stress–strain behaviour of a substance is illustrated in Figure 14.4.2. For low stresses, the graph is linear (a straight line). Within this region, the stretching is reversible. That is, the material will return to its original shape when the stress is removed. The linear section of a stress–strain graph is called the **elastic region**. Tendons are elastic, skin is elastic, even bones are elastic to a point. Once the stress becomes too large however, permanent deformation occurs. This transition stress value is called the **yield point**, or the **elastic limit**.

For higher applied forces, the stress and strain are no longer proportional, and there can be a large extension for a relatively small additional force. This is referred to as the **plastic region**. To remember this, it may help to think of stretching a plastic shopping bag—the bag will not return to its original shape. There is a maximum stress value, called the strength. Many materials, such as the one illustrated in Figure 14.4.2, fail at this maximum stress.

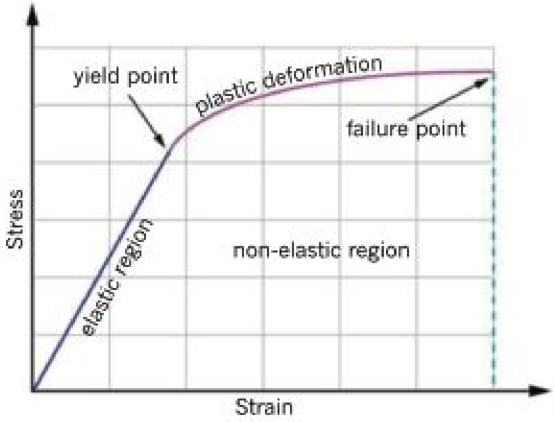


FIGURE 14.4.2 Stress–strain curve for a material, showing the elastic region, the yield point after which stress and strain are not proportional, and the material deforms plastically until it ultimately fails.

Hooke's law for a spring states that the change in length of a spring is proportional to the force that causes the deformation. For many materials, the strain is proportional to the stress causing it, for small applied forces. This follows directly from Hooke's law. Stress is simply force divided by a constant of the sample (the cross-sectional area), and strain is the extension divided by another constant (the rest length). So, if extension is proportional to force applied, strain is proportional to stress.

It is reasonable to wonder why you would want to graph stress versus strain, as opposed to graphing force versus extension, given the similarities mentioned above. The answer is that a stress–strain graph is independent of the shape of the material. As mentioned at the beginning of Section 14.3, if you apply the same force to stretch two rubber bands that are made from the same material but that have different thicknesses and lengths, they will extend by different amounts. This means that a graph of force versus extension will look different for each different shape of the same rubber material. Since stress is calculated by dividing the applied force by its cross-sectional area, and strain is calculated by dividing the extension by the original length, a stress–strain graph is not affected by the shape of the object. This means the graph of stress versus strain for a given material will be the same, regardless of the shape of the object being tested. This makes it easier to compare the properties of different materials without worrying about the influence of the object's shape. For example, Figure 14.4.3 shows the stress–strain graphs for two metals plotted on the same axes. By graphing them in this way, it is very easy to compare their properties.

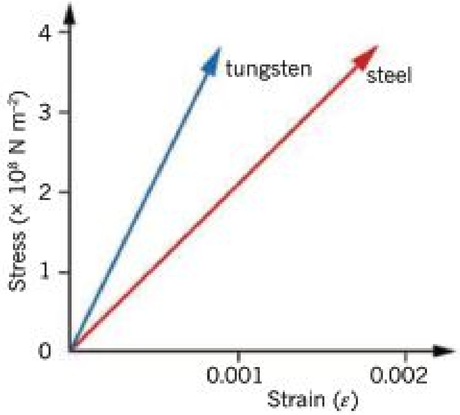


FIGURE 14.4.3 Stress and strain plotted for two different materials. These graphs allow you to compare the properties of these two materials, regardless of shape. Tungsten is stiffer than steel. Steel is more flexible than tungsten.

YOUNG'S MODULUS

The slope of the stress–strain curve in the elastic region is known as **Young's modulus** (E) for the material and is measured in N m^{-2}, or Pa. Young's modulus is a measure of the *stiffness* or rigidity of a material. It provides a direct indication of the extent of the distortion that can be expected for a given load. For example, steel and aluminium may seem like rigid materials, but Young's modulus shows that steel ($E = 2 \times 10^{11}$ N m^{-2}) is three times less yielding than aluminium ($E = 7 \times 10^{10}$ N m^{-2}). A building with an aluminium framework would bend three times farther than an identical building that used steel. Without this measure, comparing the rigidity of any two materials is difficult. Young's modulus is a property of the material and is independent of the geometry of the sample that is used to make the measurements.

Young's modulus (E) is the ratio of the stress to the strain in the linear part of the stress–strain curve:

$$E = \frac{\text{stress}}{\text{strain}} = \frac{\sigma}{\varepsilon}$$

where E is Young's modulus measured in N m^{-2} or Pa

σ is stress in N m^{-2} or Pa

ε is strain.

As an example, the Young's modulus for steel, which was stated earlier, can be calculated from Figure 14.4.3 using the gradient of the line:

$$E = \frac{4 \times 10^8}{0.002}$$
$$= 2 \times 10^{11} \text{ N m}^{-2}$$

PHYSICSFILE

Thomas Young

Thomas Young was born in Somerset, England, to Quaker parents and was the oldest of 10 children. He became an eminent physicist and physician. In 1801, the young Dr Young performed his double-slit experiment (to be studied in Unit 4) showing the wave properties of light. In his later years, Thomas Young investigated the physiology of the eye and was the first to explain astigmatism, accommodation and colour perception of the retina. He is famously known for his modulus of elasticity relating to materials. Young's modulus depends only on the particular material involved, not its dimensions. This development led to great improvements in engineering strategies.

FIGURE 14.4.4 Thomas Young (1773–1829).

Stiffness should never be confused with strength. Stiffness is measured using Young's modulus, whereas strength comes from the maximum stress that the material can endure before failure. For example, steel is stiff and strong but a wafer biscuit is stiff and weak.

Table 14.4.1 gives the Young's modulus values for some common materials as well as some biological materials. The higher the Young's modulus, the more difficult it is to deform the material. Thus rubber, at the bottom of the table, is the most yielding. Carbon fibre is light, but it stretches less than steel for the same applied force. A tendon in your body needs to stretch, while bone is stiffer to support the body, and this is reflected in their Young's moduli.

Material	Young's modulus (GPa)
cast iron	10
steel	200
aluminium	70
timber (pine) • along the grain • across the grain	10 10 1
nylon fibre	5
carbon fibre	410
marble	50
concrete	20
bone	20
tendon	1.5
human hair	10
spider silk	3
rubber	0.004

TABLE 14.4.1 Young's modulus for a variety of materials.

Biological samples such as tendons often show an initial non-linear 'toe' section in the stress–strain curve, as shown in Figure 14.4.5 on page 515. This is because the collagen fibres are crimped (compressed) when in the relaxed position. The application of a small initial force straightens the fibres, resulting in a low stiffness region. Further loading causes strain proportional to the stress applied as the fibres stretch elastically. Once the yield point is reached, damage occurs in the plastic region, after which the tendon will rupture.

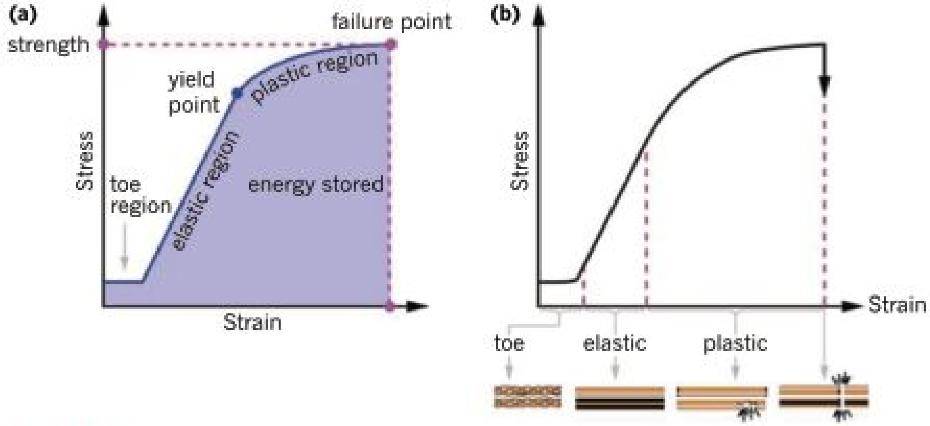


FIGURE 14.4.5 (a) Schematic plot of stress–strain curve for tendon. (b) Schematic representation of the way collagen fibres changes as strain increases in the tissue.

EXTENSION

The relationship between Young's modulus and Hooke's spring constant

There is a connection between Young's modulus, E, and the spring constant described by Hooke's law, k, since the former represents the gradient of a stress–strain graph and the latter represents the gradient of a force–extension graph. A mathematical relationship between these two constants can be derived with some manipulation of the formulas that you have learnt so far.

Rearrange the definition of Young's modulus as follows:

$$E = \frac{\sigma}{\varepsilon}$$

$$\sigma = E\varepsilon$$

Since stress is defined by $\sigma = \frac{F}{A}$, you can equate these two expressions for stress:

$$\frac{F}{A} = E\varepsilon$$

According to Hooke's law, the force applied to a spring which undergoes an extension, Δx, is given by $F = k\Delta x$. This extension is more commonly represented by Δl when describing stress and strain characteristics, so you could express Hooke's law as $F = k\Delta l$. Rewrite the above equation by substituting for F as follows:

$$\frac{k\Delta l}{A} = E\varepsilon$$

Rearranging this equation gives:

$$k = EA\left(\frac{\varepsilon}{\Delta l}\right)$$

Strain is defined as $\varepsilon = \frac{\Delta l}{l}$, which can be rearranged to $\frac{\varepsilon}{\Delta l} = \frac{1}{l}$.

This means that the spring constant, k, can be expressed in terms of Young's modulus, E, as follows:

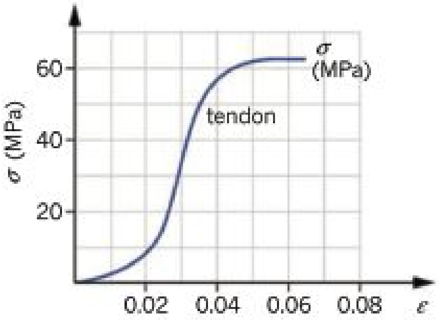
$$k = E\frac{A}{l}$$

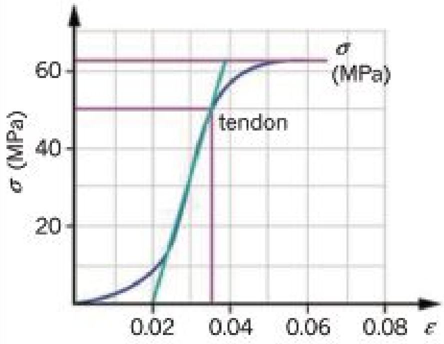
In words, the spring constant is equal to the Young's modulus multiplied by the ratio of the cross-sectional area of the object being analysed, divided by its length. So the difference between the two constants is purely to do with the geometry of the object, which is representative of the difference between a force–extension graph (which depends on the shape of an object) and a stress–strain graph (which describes material properties only, and is independent of the shape of the object).

Worked example 14.4.1

CALCULATING YOUNG'S MODULUS

Estimate the Young's modulus of the tendon in the elastic region using the graph below, and also find the strength of the tendon:

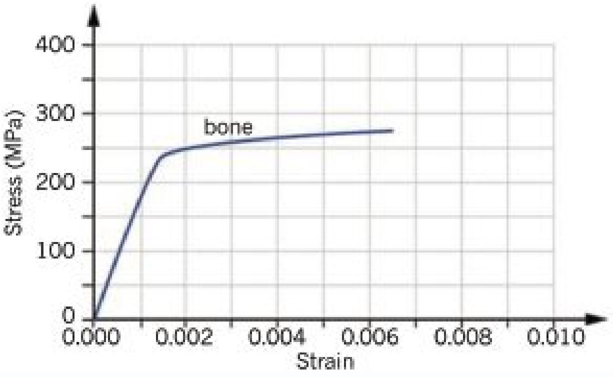


Thinking	Working
Firstly find the elastic region.	It is the linear portion of the graph.
The Young's modulus is the slope of the linear section of the graph. Choose two points on a line which passes through the linear section for calculating the gradient.	
Calculate Young's modulus. Note that stress is measured in MPa on this graph.	$E = \frac{\text{change in stress}}{\text{change in strain}}$ $= \frac{(50 - 0)\text{ MPa}}{(0.036 - 0.020)}$ $= 3.1\text{ GPa}$
The strength of the tendon is the maximum stress it can withstand before failure.	The strength is read from the graph on the stress axis. strength = approximately 63 MPa

Worked example: Try yourself 14.4.1

CALCULATING YOUNG'S MODULUS

Find the Young's modulus of bone from the graph below and also find the strength of bone (give your answer correct to two significant figures).



Human skin is elastic, as it can stretch and then return to its initial configuration. This is because of the arrangement of elastin fibres illustrated in Figure 14.4.6(a). Collagen fibres are interspersed with elastin in the skin to add strength. Young skin is elastic, but over time the collagen and elastin fibres deteriorate, and plastic deformation occurs. The skin stretches but does not recover. The man in Figure 14.4.6(b) suffers from a connective tissue disorder that interferes with collagen production, and so his skin is able to stretch excessively.

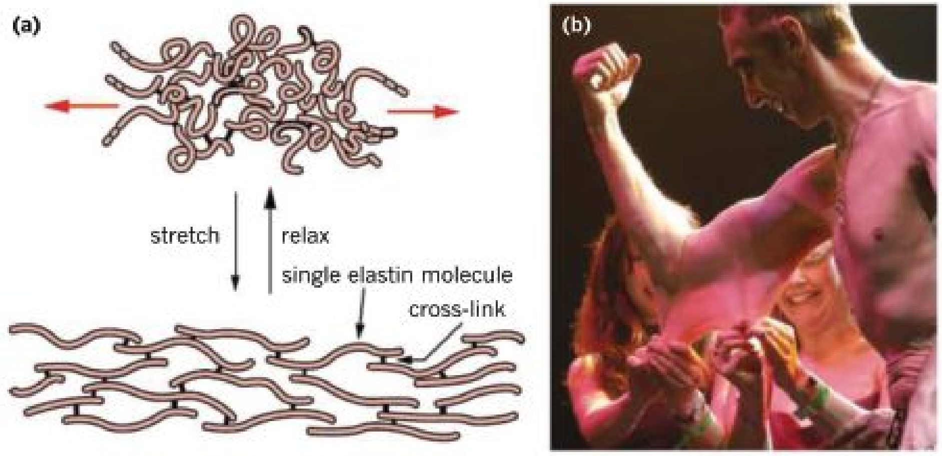


FIGURE 14.4.6 (a) Arrangement of elastin fibres in skin, showing how the skin can stretch and relax elastically. (b) Collagen gives stiffness and strength to the skin. This man has a connective tissue disorder that interferes with collagen production, leaving his skin exceptionally stretchy.

MORE INFORMATION FROM STRESS–STRAIN CURVES

While many materials show an elastic region in the stress–strain curve, they differ significantly in their response to high stress. **Brittle** materials fail abruptly with no plastic deformation. The plate in Figure 14.4.7, for example, is made from a brittle material. As people age, a condition known as osteoporosis makes their bones become more porous and brittle. Fractures arising from a fall are relatively common in the elderly, while the bones of young people are more flexible.

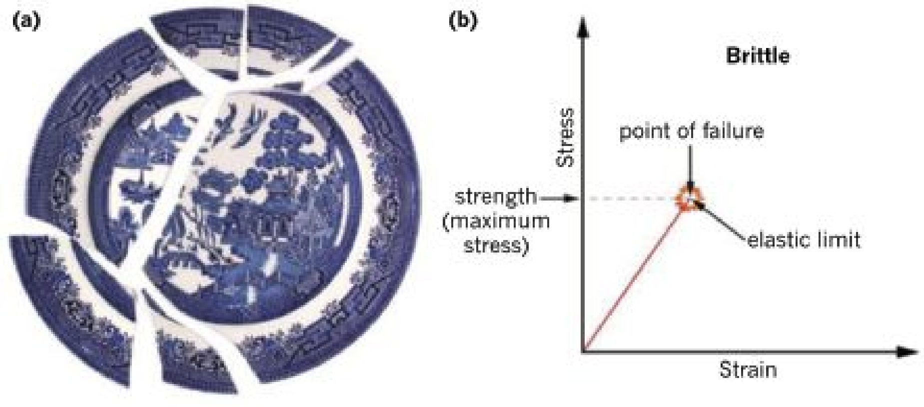


FIGURE 14.4.7 (a) This china plate is brittle. It has fractured without deforming at all. (b) The stress–strain graph for the plate shows that it exhibited no plastic behaviour. Its maximum stress (strength) was also its elastic limit.

PHYSICSFILE

Brittle versus weak

It is important to note that brittle materials are not necessarily weak. For example, some ceramics, which are brittle materials, can have enormous strength. The ceramic tiles used on the space shuttle have a breaking stress much greater than that of most metals.

Metals are known to be **ductile,** meaning that they can be drawn into wires. Once the yield point is reached, the sample will thin and become a weak point. This deformation is called necking. The metal then continues to stretch significantly before finally failing. It is common for the maximum stress value to lie in the plastic region, and for the material to finally fail at a lower stress value than this maximum (see Figure 14.4.8).

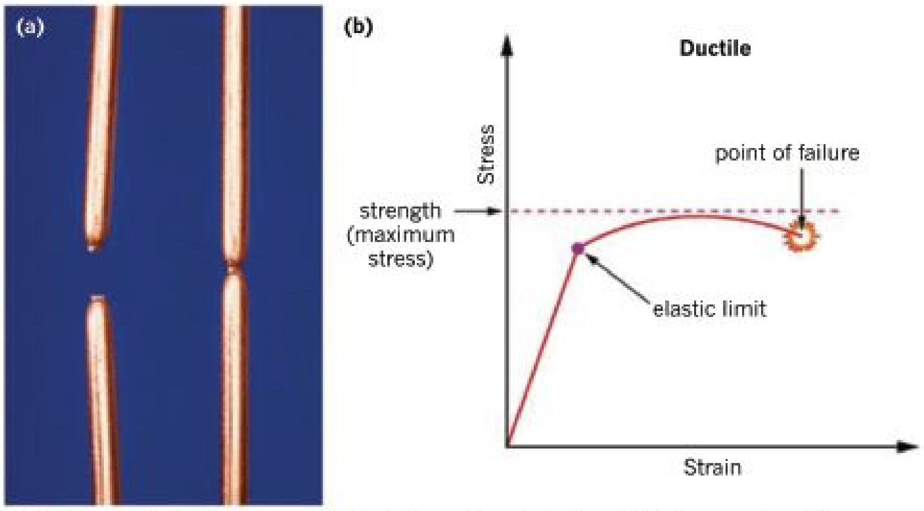


FIGURE 14.4.8 (a) As the copper wire is stretched beyond its elastic limit, one section will narrow and become a weak point. The wire will continue to stretch, then snap or fracture at this weak point. (b) The stress–strain curve for a ductile material shows a large plastic region. The stress at which the copper wire fails is well below the maximum stress that it has withstood.

Most biological samples show a property known as **viscoelasticity**. Viscous materials have a time-dependent response to stress. Thus the stiffness of tissues depends on the rate at which the force is applied. Also, when the material is deformed, the return to its original shape is time-delayed.

PHYSICSFILE

Viscoelasticity

It is easy to demonstrate viscoelasticity with the stretching of a hair sample. A protein called keratin exists in two forms in human hair. The α-keratin is arranged in helix formation held in place by hydrogen bonds, while τ-keratin molecules lie in sheets. Applying stress to the hair results in the straightening out of the α-keratin to form the linear τ-keratin. When the stress is removed, the helices reform over time and the hair recovers its original length.

STRAIN ENERGY

Much of the discussion so far has concentrated on the response of materials to a force. In physics, all mechanical situations can also be analysed from the point of view of energy.

When studying Hooke's law (Section 12.2), the energy stored by a spring was calculated from the area under the force–extension curve.

Work done is defined as the product of force, p. 428, applied and distance moved in the direction of the force. For the stress–strain curve, a force causes a deformation Δl in the direction of the force, and so work is done.

Consider that the area under the stress–strain curve is $\sigma\varepsilon$. This may be written as $\frac{F}{A} \times \frac{\Delta l}{l}$ or $\frac{F\Delta l}{Al}$. Since $F\Delta l$ is the work done, and Al is the volume of the material, the area under the stress–strain curve is the work done per unit volume in loading the material. The units are J m^{-3}.

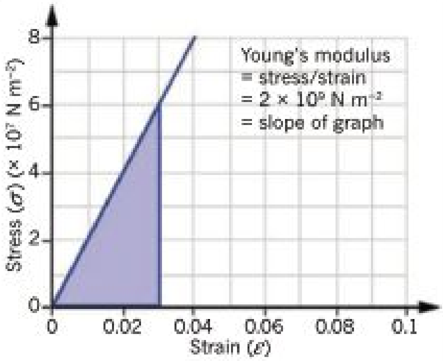
The **strain energy** is the work done in changing the length of a material, and the strain energy per unit volume can be found from the area under a stress–strain curve. Within the elastic region of a stress–strain graph, the strain energy per unit volume will be given by $U = \frac{\sigma\varepsilon}{2}$. Using the relationship between Young's modulus, stress and strain, the strain energy per unit volume can be expressed as $U = \frac{E\varepsilon^2}{2}$.

For an elastic deformation, the work done in stretching or compressing the material is recovered when the material relaxes. While tendons are not perfectly elastic, they store energy when they stretch, and return a large percentage of the energy as they relax. Kangaroos benefit from the high energy efficiency of the tendons in their legs and can maintain their hopping over large distances as a result.

Worked example 14.4.2

CALCULATION OF ENERGY PER UNIT VOLUME STORED IN A MATERIAL

A kangaroo tendon is found to have Young's modulus 2.0×10^9 Pa. In normal operation, the tendon stretches by 3% of its rest length. Calculate the energy that is stored in the tendon per cubic metre.

Thinking	Working
Convert the strain value to a decimal, rather than a percentage.	$\varepsilon = 3\%$ $= 0.03$
Remember that the strain energy is the area under a stress–strain curve. Make a quick sketch of the curve, calculating the stress from $\sigma = E\varepsilon$.	 $\sigma = E\varepsilon$ $= 2.0 \times 10^9 \text{ Pa} \times 0.03$ $= 6 \times 10^7$ Pa
The strain energy per cubic metre is the area under the curve: $U = \frac{1}{2}\sigma\varepsilon$	$U = \frac{1}{2}\sigma\varepsilon$ $= \frac{1}{2}(6 \times 10^7)(0.03)$ J $= 0.9 \times 10^6 \text{ J m}^{-3}$ $= 0.9 \text{ MJ m}^{-3}$

Worked example: Try yourself 14.4.2

CALCULATION OF ENERGY PER UNIT VOLUME STORED IN A MATERIAL

When the kangaroo hops more strenuously, the tendon stretches by 5% of its rest length. Calculate the strain energy per m^3.

Whenever a material deforms irreversibly, some energy is transformed in the physical change that takes place, and this energy is not recovered from the sample when the load is removed. Usually this energy is dissipated as heat. This means the sample will behave differently as a result of the load it has sustained. This is called **hysteresis**, which is a memory effect where the properties of a material depend on its prior treatment. The stress–strain curve for the material as it is loaded is not the same as when it is unloaded. This is shown in Figure 14.4.9. The load curve is different from the curve showing the material as it is unloaded. The area under the loading curve is the work done in loading the material, while the area under the unloading curve is the energy returned when the material relaxes. The area contained within the hysteresis loop represents the energy that is dissipated in the material, usually as heat.

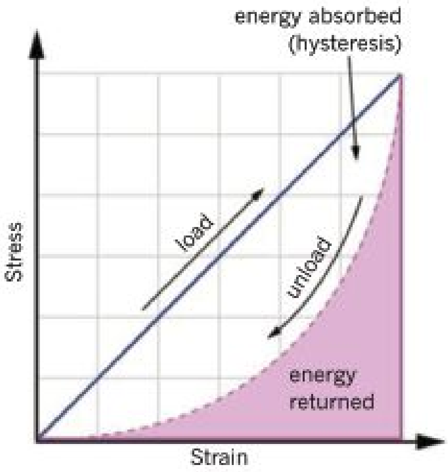


FIGURE 14.4.9 Stress–strain curve for an ideal viscoelastic material.

PHYSICSFILE

Stress–strain behaviour in rubber bands

A simple rubber band does not exhibit simple stress–strain behaviour. Rubber bands do not follow Hooke's law. They behave in a non-linear manner as they are stretched and as they are unloaded, before finishing up with close to the original dimensions. The unloading curve is different from the loading curve, indicating that some energy is dissipated during this process. If you stretch and unstretch a rubber band a number of times, you should be able to feel this as heat that is lost.

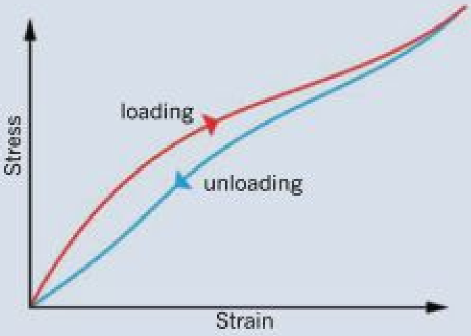


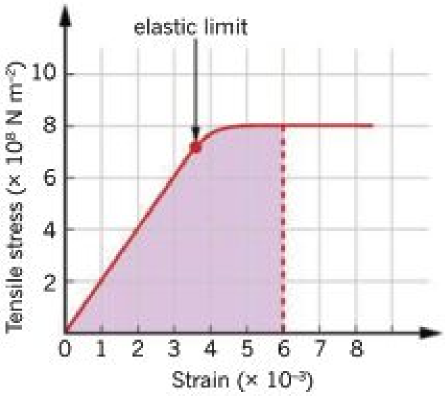
FIGURE 14.4.10 The hysteresis curve displayed by a rubber band indicates that it has different loading and unloading behaviour. The amount of energy that is transformed into heat is indicated by the size of the area between the lines.

Worked example 14.4.3

CALCULATING STRAIN ENERGY

A material that is being tested for its suitability in hip replacements has the stress–strain curve shown. A 20 cm sample with cross-sectional area 5 cm^2 is stressed beyond its yield point to a strain of 6×10^{-3}.

Calculate the strain energy stored in the material under stress.



Thinking	Working
Recall how to calculate strain energy per unit volume from a stress–strain curve.	The strain energy per unit volume is given by the area under the stress–strain curve.
Find the energy per unit volume represented by each square.	Each square represents $(2.0 \times 10^8)(1.0 \times 10^{-3})$ $= 2.0 \times 10^5$ J m^{-3}
The total strain energy per unit volume is found by counting squares in this case.	Strain energy per unit volume = 16 squares × 2.0×10^5 J m^{-3} $= 3.2 \times 10^6$ J m^{-3}
To find the actual energy stored in the sample, the strain energy per unit volume is multiplied by the actual volume. Remember to convert centimetres to metres. 1 cm = 10^{-2} m 1 cm^2 = 10^{-4} m^2	Energy = strain energy per unit volume × volume $= 3.2 \times 10^6$ J m^{-3} × $(0.20)(5.0)(10^{-4})$ m^3 $= 3.2 \times 10^2$ J

Worked example: Try yourself 14.4.3

CALCULATING STRAIN ENERGY

An identical sample (20 cm × 5 cm^2) of the same material was stressed until it fractured at a strain of 8.5×10^{-3}. Use the stress–strain curve from Worked example 14.4.3 to calculate the total energy needed to fracture the sample.

The **toughness** of a material is its ability to absorb energy while experiencing plastic deformation. This energy can be estimated by finding the strain energy for the material up to the point of failure, as was done in Worked example: Try yourself 14.4.3. The larger the area under the stress–strain curve, measured in J m^{-3}, the more energy the material will absorb before failing. Brittle materials are generally not tough, while substances that can undergo large plastic deformations are tougher. The graph in Figure 14.4.11 shows two different materials—one tough and one brittle. Material X is not very tough. While it is stiffer and stronger than material Y, not much energy is needed to break it, as indicated by the small area under the graph. By way of contrast, material Y is very tough. A large amount of strain energy (as indicated by the large graph area) is required before it fails. Y undergoes a lot of plastic deformation before breaking.

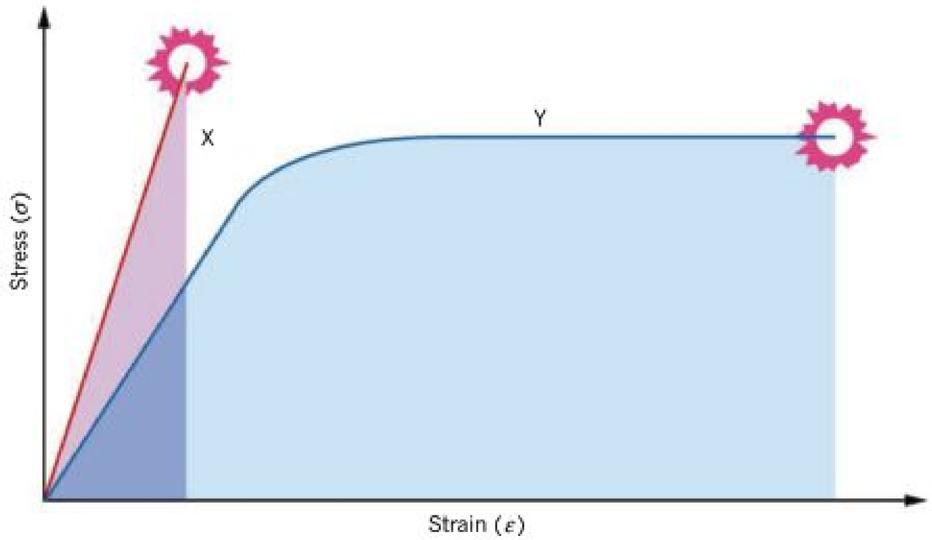


FIGURE 14.4.11 The stress–strain graph of two different materials, X and Y. Material X is not as tough as material Y, even though it is stiffer and stronger.

14.4 Review

SUMMARY

- When the strain is proportional to the applied stress, the stress–strain curve is linear and the material is said to behave elastically.
- The Young's modulus is the gradient of the stress–strain curve. It is an indication of the stiffness of the material, and is defined by the equation $E = \frac{\text{stress}}{\text{strain}} = \frac{\sigma}{\varepsilon}$ where E is measured in N m^{-2} or Pa.
- In the elastic region, the material regains its original dimensions when the stress is removed. In the ideal case, all the energy stored while the material was deformed is returned.
- The strength of a material is the maximum stress it can withstand before failing.
- Beyond the yield point on the stress–strain curve, the material deforms permanently. Relatively small additional stress will cause significant deformation. This is called plastic behaviour.
- Brittle materials do not have a plastic region and fail abruptly at the yield point.
- Ductile materials such as metals have a large plastic region, indicating that they can be drawn into wires.
- Biological materials are often viscoelastic, which means that the stress–strain behaviour is time-dependent, with the stiffness depending on the rate at which a force is applied.
- The work done per unit volume in deforming the material is called the strain energy per unit volume, and it may be calculated from the area under a stress–strain curve.
- Many biological samples show hysteresis. The stress–strain loading curve is not the same as the unloading curve. This also means that not all the energy is recovered when the sample relaxes.
- The toughness of a material is its ability to absorb energy while experiencing plastic deformation. The toughness is the strain energy up to the point of failure.

KEY QUESTIONS

1 Collagen is the main protein of human connective tissues. What is the strength of a collagen fibre?

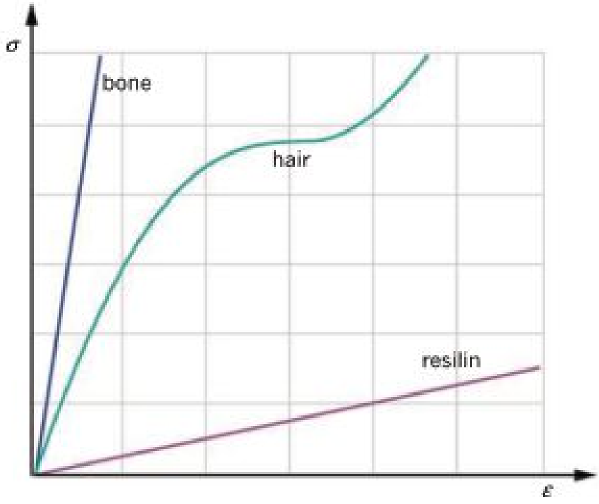
A the maximum stress that a fibre can sustain before it fails

B the maximum force that can be applied to a fibre before it fails

C the strain that a fibre can sustain before it fails

D how easily the fibre breaks

2 Three biological materials are presented in the stress–strain graph below. Rank each of the materials from highest to lowest for the following properties:



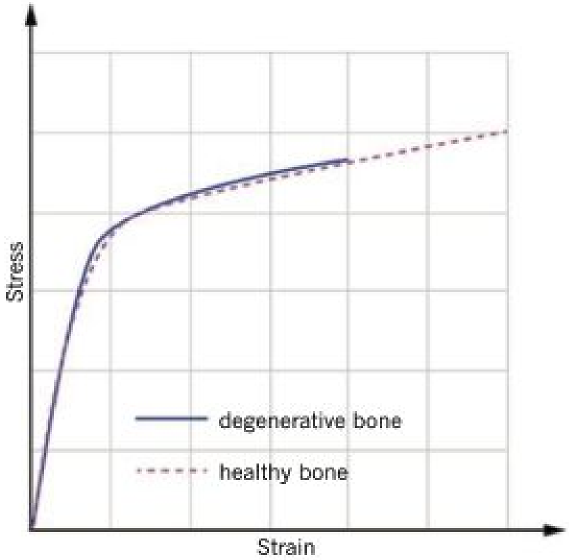
a strength

b stiffness

c toughness

d plasticity.

3 It is well known that as people age their bones tend to degenerate. The graph below shows a stress–strain curve for the bones of young people and elderly people. Choose the response(s) that best correspond to the data given.



A The degenerative bone is significantly stiffer than the healthy bone.
B The degenerative bone can only withstand half of the stress of the young bone.
C The healthy bone can withstand greater stress and strain.
D The degenerative bone can withstand the same strain as the young bone.
E The strength of the healthy bone is significantly greater than that of the degenerative bone.
F The elasticity of both bone types is the same.
G Healthy bone has a greater plastic range before the bone fails.

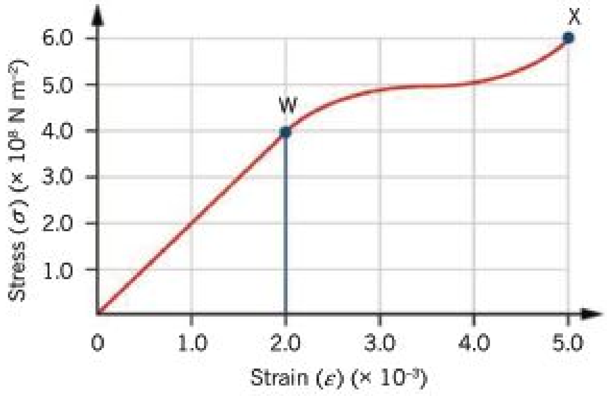
The following information is relevant to questions 4 and 5. A 10 cm length of tendon 10 mm² in cross-section is subjected to forces by hanging weights from the sample. The length of the tendon is measured for each weight, and the data is used to create the following table.

Initial length (cm)	Final length (cm)	Δ*l* (cm)	Strain	Mass (kg)	Force (× 10² N)	Stress (MPa)
10.0	0	0	0	0	0	0
10.0	10.25	0.25	0.025	25	2.5	25
10.0	10.50	0.50	0.050	50	5.0	50
10.0	10.75	0.75	0.075	75	7.5	75
10.0	11.00	1.00	0.10	100	10	100
10.0	12.00	2.00	0.20	110	11	110

4 Draw a stress–strain graph for the tendon.

5 Use your graph to calculate the Young's modulus of the tendon.

The following information applies to questions 6–12. Steel is being tested for its suitability to be used in prostheses. The following graph shows the stress–strain relationship for the particular steel alloy used. At point W the wire loses its elastic properties, and it snaps at point X.



6 **a** What is the maximum stress that this wire can tolerate before it undergoes plastic deformation?
b What is the largest strain that this wire will tolerate while still obeying Hooke's law?
c Calculate the gradient of this graph in the interval up to and including the elastic limit.
d What physical constant does the gradient represent?
e What is the tensile strength of this material?

7 A 1.0 m piece of this steel wire has a radius of 1.0 mm.
a What is the maximum extension that this wire can tolerate without breaking?
b What is the maximum extension that this wire can tolerate before it undergoes plastic deformation?

8 A tensile force of 1.0 kN is applied to the wire described in question 7. Use the graph above to find the subsequent extension of the wire:
a while the force is still acting
b some time after the force has been removed.

9 This wire is now placed under a tensile stress of 5.0×10^8 Pa.
a Up to this point, has the wire exhibited brittle or ductile behaviour?
b What is the extension of the wire while it is supporting this load?
c If the load is removed, will the wire return to its original length? Explain your answer.

10 A 4.0×10^{-5} m³ volume of this steel alloy is subjected to a stress such that it undergoes a strain of 0.0020.
a Calculate the work done per cubic metre to produce this strain.
b Calculate the strain energy stored in the alloy at this point.
c Describe how the alloy will behave when the stress is removed. Will it resume its original length? Will the alloy heat up?

11 The stress on the 4.0×10^{-5} m³ volume of this steel alloy is increased so that the strain experienced by the alloy increases to 0.0040.
a Calculate the work done per cubic metre to produce this strain.
b Calculate the strain energy stored in the alloy at this point.
c Describe how the alloy will behave when the stress is removed. Will it resume its original length? Will the alloy heat up?

12 Finally, a stress of 6.0×10^8 N m^{-2} is applied to the 4.0×10^{-5} m³ volume of alloy.
a What happens to the alloy at this point?
b Calculate the strain energy per cubic metre that was needed to fracture the alloy.
c Calculate the strain energy needed to cause the alloy to fail.

14.5 The future: Materials for use in prosthetics

FIGURE 14.5.1 A prosthetic leg.

A thorough understanding of the properties of biological materials allows medical personnel to understand the cause and treatment of various injuries. When tissues fail altogether, materials science once again comes into play when biomedical scientists look for materials to manufacture replacement parts. Some prostheses, such as artificial limbs for amputees (Figure 14.5.1), are external. Others are internal, as in hip or heart valve replacements. Either way, the mechanical properties of the materials used are of crucial importance if they are to effectively replace biological tissues.

REPLACING BONE: PROSTHETIC LIMBS AND HIP REPLACEMENTS

In materials used to replace bone, strength, stiffness, elasticity and durability are all important. Modern prosthetic limbs are frequently made of carbon fibre composites that are lightweight, stiff and strong.

Composite materials

Composite materials are composed of more than one substance, employing favourable properties of each. Carbon fibre itself has tensile strength superior to steel, but an unsupported fibre has no compressive strength at all. The fibre is generally implanted in a material such as a polymer or epoxy resin. In this arrangement, the polymer or epoxy resin is called the binder material, or the matrix. The matrix is there to keep the fibres in position, and to prevent buckling of the fibres by providing compressive strength. As a result, the composite material has excellent tensile and compressive properties.

Reinforced composites are generally **anisotropic**, meaning that their properties are not uniform in all directions. Figure 14.5.2 illustrates how a reinforced polymer or epoxy would only be strong along the axis aligned with the fibres. This is because the fibres are not able to provide tensile strength in the perpendicular direction, and the matrix material is generally not very strong under tension. In **laminates**, many layers are applied on top of each other in different directions. This gives the material the desired properties because there will always be some layers of the material in which the fibres align with the direction of the stress. Poor placement of the fibres can leave the mechanical properties of the material dominated by the properties of the matrix material, and the expensive fibres having little effect.

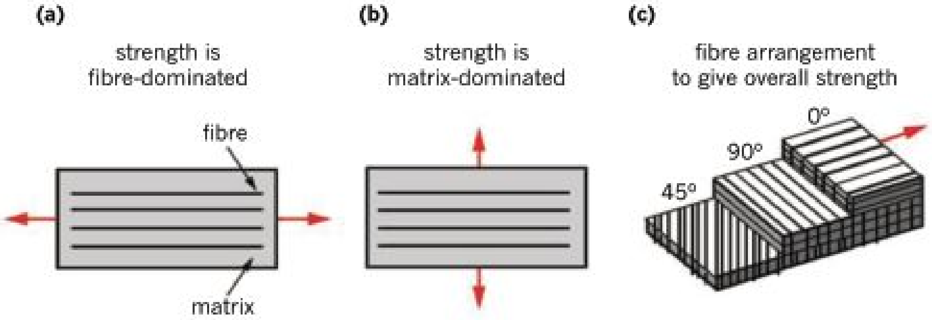


FIGURE 14.5.2 (a) A composite material is strong when pulled parallel to the reinforcing fibres. (b) Material strength perpendicular to the fibres is dominated by the mechanical properties of the matrix. (c) Laminates are made from multiple layers of fibres.

Prosthetic limbs

As a result of the significant advances in materials science in recent years, prosthetic limbs can be designed with particular properties to serve certain purposes extremely well. The running blades that have become commonplace in Paralympic athletics events are made from carbon fibre laminates (see Figure 14.5.3). They are finely engineered to function extremely efficiently in running races. In particular, they must have sufficient strength to withstand the forces produced while the runner is accelerating, and a large enough elastic region to allow significant flexing so that kinetic energy can be stored as strain energy when compressed, and then returned to kinetic energy. The prostheses have been so effective that there has even been debate as to whether they provide an advantage to amputees over able-bodied athletes.

FIGURE 14.5.3 Single and double amputees wearing carbon fibre laminate running blades.

Hip replacements

Internal prostheses place even more stringent requirements on a material because they need to be non-toxic and resistant to corrosion, and not be rejected by the body. In other words, they need to be biocompatible with the human body. In addition, because an operation is required for implantation, the materials need to be durable over time.

One of the most common orthopaedic operations on elderly people is a hip replacement. This requires that a stem and ball be attached to the top of the femur, as shown in Figure 14.5.4. The ball fits into a cup that replaces the hip socket in the pelvis. The shaft is generally titanium or a chromium and cobalt alloy. The ball can be metal or ceramic and the cup is metal, ceramic or polyethylene.

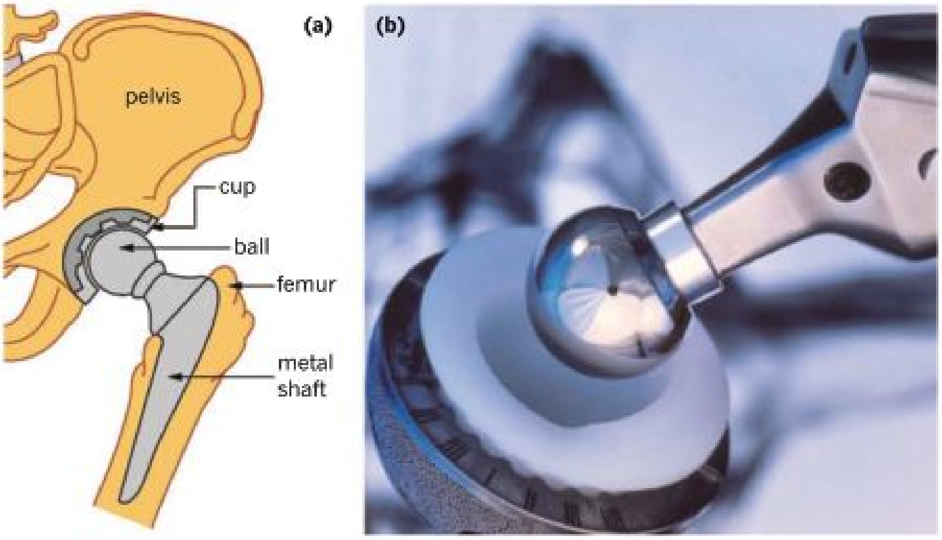


FIGURE 14.5.4 (a) Diagram of prosthetic implant for hip replacement. (b) A prosthetic hip ball-and-socket joint.

Friction inevitably results in some wear on the joint and, particularly when replacements are carried out on relatively young patients, the durability of prosthetics is a concern. If the body recognises the materials as foreign bodies, an immune response can be triggered, with very harmful consequences for the patient. In addition, the bone can degrade, resulting in loosening of the prosthesis itself.

REPLACING CONNECTIVE TISSUE: HEART VALVES

A typical heart valve opens and closes 40 million times a year. The material making up the valve has to be flexible enough for the blood itself to be able to push open the valve during systole (contraction of the heart), but rigid enough to support the weight of the blood above it and avoid backflow during diastole (contraction of the heart) when the valve is closed.

The heart valve leaflets are composed of three layers, each of which has specific functions. A layer rich in collagen fibres (a type of protein) gives mechanical strength to the leaflets on the aortic side. These fibres are strong under tension, and are arranged circumferentially, as shown in Figure 14.5.5. The wavy collagen fibres are able to straighten, which allows the valve initially to stretch under relatively small loads. Once straightened, the fibres become significantly stiffer under the higher load imposed by diastole.

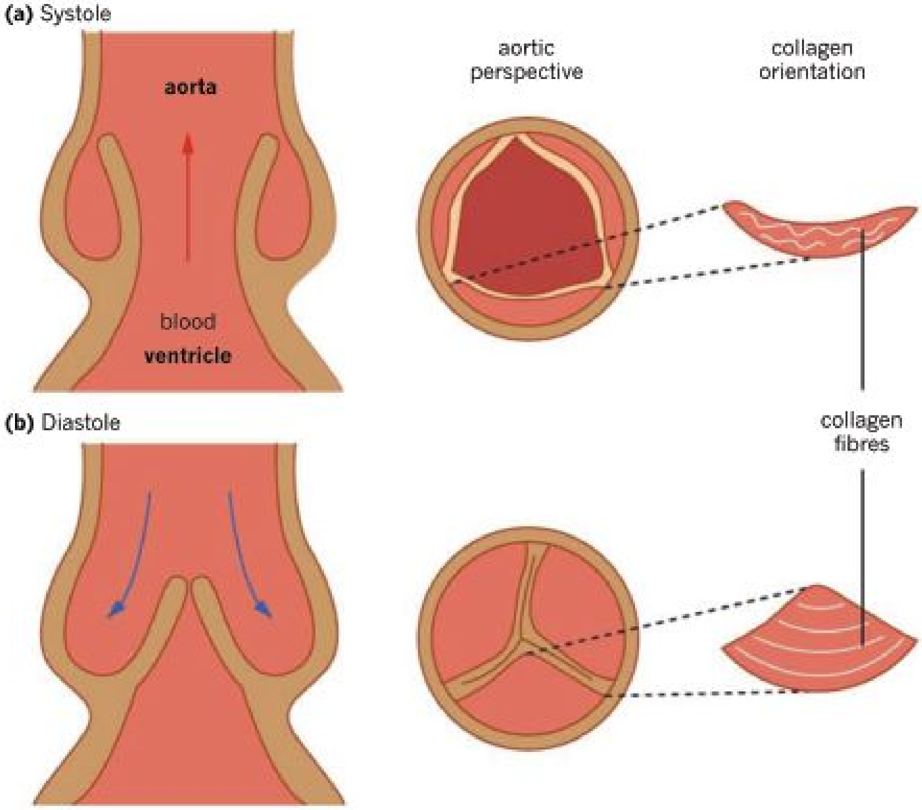


FIGURE 14.5.5 (a) During systole, the blood pressure pushes the valve open and the collagen leaflets are relaxed and lie in a crimped position. (b) During diastole, the leaflets close the valve and the circumferentially aligned collagen fibres provide tensile strength, which supports the weight of the blood above the valve.

The middle layer of a heart valve leaflet is a lubricating and shock-absorbing layer, and the layer closest to the ventricle is high in elastin that is elastic and not as stiff as the collagen layer. The elastin helps the valves close effectively.

Heart valves can be replaced with mechanical valves, and with biological surrogates such as valves made from animal tissue. A range of different valves is shown in Figure 14.5.6. Mechanical valves require that the patient take blood-thinning medication to prevent clot formation. The valves are more durable than animal prosthetic valves and are thus preferred for younger patients. Valves from animal tissue may more closely mimic the flow of blood in a normal valve, but are not as durable and so are preferred in elderly patients whose life expectancy is such that a replacement is unlikely to be required.

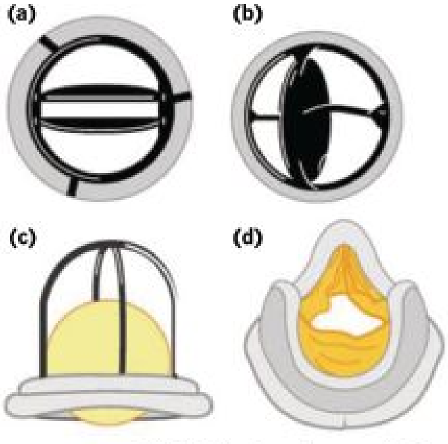


FIGURE 14.5.6 (a) Bi-leaf valve, (b) mono-leaf valve, (c) ball-and-cage valve, (d) bioprosthetic valve made from animal tissue. Prosthetic heart valves (a)–(c) are mechanical valves.

THE CHALLENGES

Biological tissues are very complex, so finding an artificial material that mimics the original tissue is challenging. The body is sensitive to invasion by foreign objects, and so the interaction of the prosthesis and the native tissue has to be carefully managed. Even when it is possible to find a material with suitable properties, and one that will be compatible with other tissues, the durability and maintenance of the prosthesis can be a challenge.

In countries such as Australia, an ageing population points towards a growing need for prostheses. There is much work to be done to model biological tissues accurately, and to design and test new materials, and the physics principles introduced in this chapter form the foundation for this research.

14.5 Review

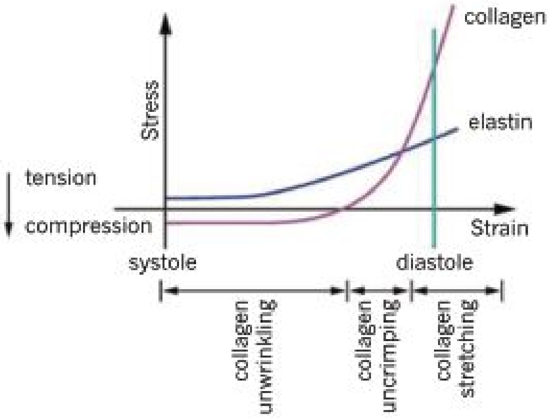
SUMMARY

- Materials used in prostheses need to match the material properties of the biological tissues that they replace.
- Practical issues such as cost and durability are also important factors.
- External prostheses such as limbs for amputees are often made from carbon fibre laminates.
- When fibres are embedded in a matrix, the composite is strong under tension and compression.
- Internal prostheses place further constraints on the materials that can be used because they need to be non-toxic, they must not corrode in situ and neither should they allow clot formation or encourage bacterial growth.
- Heart valves contain collagen for strength and elastin for elasticity.
- Replacement valves may be mechanical, or made of animal tissues that have similar properties to the original valve tissue.

KEY QUESTIONS

1 The following factors need to be considered when choosing materials for external prostheses. Give reasons why each factor is important.
  - **a** durability
  - **b** strength
  - **c** flexibility
  - **d** stiffness
  - **e** high strength-to-mass ratio
  - **f** elasticity

2 The diagram shows the stress versus strain characteristics for two of the most important connective tissues in heart valves: elastin and collagen. The loading conditions of systole (open valve) and diastole (closed valve) are indicated by labelled vertical lines. The loading conditions for straightening and uncrimping of collagen fibres are shown below the *x*-axis. Answer the following questions based on the diagram.



  - **a** Compare the stiffness of collagen and elastin during systole and diastole.
  - **b** Which of the two materials is more flexible?
  - **c** Which material offers more strength during diastole?

3 Which of the following statements with regard to prosthetic materials is false?
  - **A** Composites that are made of fibres embedded in a matrix are better suited to resist tension and compression.
  - **B** Internal prostheses need to be biocompatible.
  - **C** Carbon fibre laminates are often used for amputee limb-replacement prostheses.
  - **D** Materials used for heart valve replacement need to be strong but flexible.
  - **E** None of the above.

4 Of the following, which one is an undesirable feature when choosing a material for heart valve replacement?
  - **A** The material is biocompatible, i.e. it displays low toxicity.
  - **B** The material is resistant to corrosion.
  - **C** The material does not promote clot formation or bacterial growth within tissues.
  - **D** The material promotes growth of different types of cells, including those of bacterial and fungal origin.

5 Do you think it is more difficult to find a suitable material for an internal or an external prosthesis? Justify your answer.

Chapter review

14

KEY TERMS

anisotropic
brittle
composite material
compressive force
ductile
effort
elastic limit
elastic region
force multiplier
gravitational force
hysteresis
laminate
lever
load
mechanical advantage
normal reaction force
plastic region
prostheses
shear force
speed multiplier
strain
strain energy
strength
stress
tensile forces
toughness
translation
viscoelasticity
yield point
Young's modulus

1 Select the incorrect statement.
- **A** The gravitational force causes the compression of load-bearing bones.
- **B** The load-bearing bones of an astronaut in the International Space Station are generally under compression.
- **C** Tendons operate under tension caused by muscular contraction.
- **D** The discs in the spinal column are under compression.

2 Select the person or people who are not in translational equilibrium under the following circumstances (more than one answer is possible).
- **A** a boy leaving the surface of a trampoline and travelling upwards
- **B** a girl swinging to and fro on a swing
- **C** a passenger travelling at constant speed on a straight road
- **D** an ice skater spinning rapidly on the tip of her ice skate

3 Why do tendons have large tensile strength and toughness, yet have minimal compressive strength?

4 Which of the following terms describes the amount of energy that a material can absorb per unit volume before failing: plasticity, toughness, strain or energy?

5 In which type of material is the yield point the same as the point of failure?

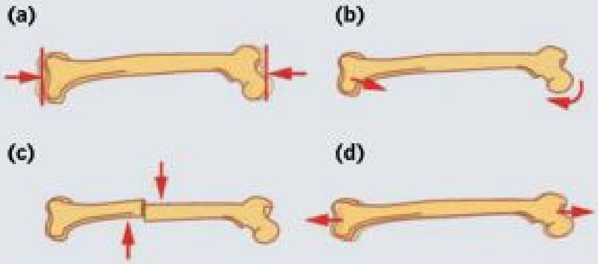
6 Choose the true statement(s) for plastic materials.
- **A** The strain energy to the yield point is greater than the toughness.
- **B** The toughness is greater than the elastic energy to yield point.
- **C** The strain does not increase linearly with stress before the yield point.
- **D** The strain does not increase linearly with stress after the yield point.

7 Choose the true statement(s) for perfectly elastic materials.
- **A** They have linear stress–strain curves.
- **B** They do not absorb energy when they are stressed and then relax.
- **C** They are very tough.
- **D** They are ductile.

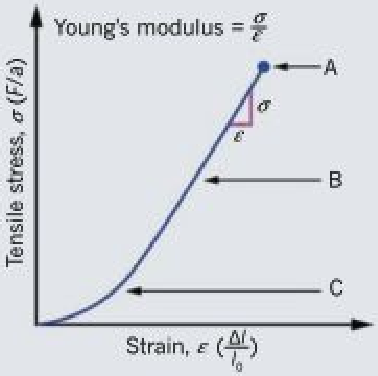
8 Choose the true statement(s) for viscoelastic materials.
- **A** They absorb energy when they are stressed and then relax.
- **B** They show hysteresis.
- **C** They are ductile.
- **D** They have stiffness that depends on the rate of loading.

9 Choose the true statement(s) regarding a plastic material undergoing a loading and unloading cycle.
- **A** The material absorbs energy.
- **B** Energy is transformed as the material structure is altered.
- **C** The energy stored in the loading is more than the energy returned when the material is unloaded.
- **D** All the energy that was stored in loading is released on unloading.
- **E** Energy is transformed into heat.
- **F** The energy stored in the loading is less than the energy returned when the material is unloaded.
- **G** Irreversible structural changes occur.

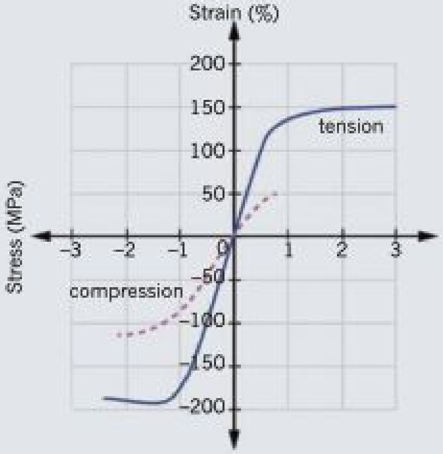
10 Identify the types of stresses experienced by the femur in each diagram shown.



11 Consider the stress–strain curve for collagen fibres shown below. Explain what is happening to the collagen fibres at each of the points A, B and C indicated on the graph.



12 The graph shows data collected for load-bearing bone under tension and compression. The positive stresses represent tension and the negative stresses, compression, as marked. The blue curve is for bone tested along the longitudinal direction and the purple curve is for forces directed in the transverse direction—that is, across the bone.

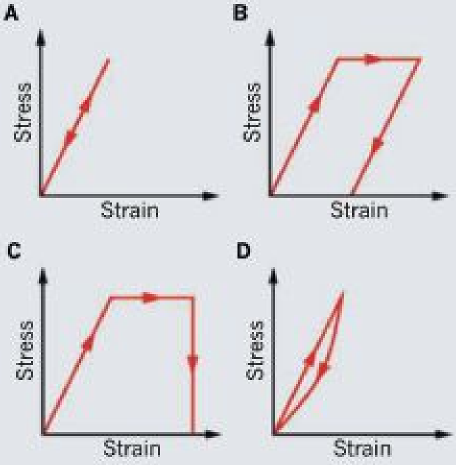


Use the data from this graph to decide whether the following descriptions apply to tests in the longitudinal direction or in the transverse direction.

- **a** strongest under tension
- **b** strongest under compression
- **c** brittle under tension
- **d** toughest under tension
- **e** toughest under compression
- **f** stiffest under tension
- **g** behaves plastically under high tensile and compressive forces.

13 Using the data from question 12, comment on the statement that bone is anisotropic, and that the structure of bone is adapted to its function. Propose a scenario when the inferior properties in the transverse direction could be a disadvantage.

14 Answer the questions that follow concerning the stress–strain curves for the loading and unloading of each material.



- **a** Which material(s) show perfect elastic behaviour?
- **b** Which material(s) show some plastic behaviour?
- **c** Which material(s) have residual strain after unloading?
- **d** Which material(s) do not absorb energy in a complete loading/unloading cycle?

15 A piece of dried horse tendon has a Young's modulus of 890 MPa, and it fails at 9.0% strain. Assuming perfect elastic behaviour, calculate the tensile strength of the tendon and plot its stress–strain curve.

16 Using the data from question 15, calculate the toughness of the tendon.

17 Correct fitting of a backpack requires that the internal frame is adjusted so that the pack fits snugly against the back, chest straps should be fastened, and the hip belt is firmly fastened. Heavier objects are packed closer to the back with lighter objects further from the body

Explain the value of the features described above using what you have learnt in this chapter.

18 Carbon fibre composites are often used for prostheses because of their high strength under both tension and compression. Explain:

a what provides the tensile strength for carbon fibre composites

b what provides the compressive strength for carbon fibre composites

c how to achieve high tensile strength in multiple directions for a carbon fibre composite.

19 The value of Young's modulus for aluminium is 7.0×10^{10} N m^{-2}. An aluminium rod of radius 5.00 mm used in an external prostheses is subjected to a compressive force during which its length is decreased by 1.0%. Calculate the value of the force acting on the rod.

The following data for human bone apply to questions 20 and 21.

Tensile strength = 1.2×10^8 N m^{-2}
Compressive strength = 1.7×10^8 N m^{-2}
Young's modulus = 1.6×10^{10} N m^{-2}
Elastic limit = 1.0×10^8 N m^{-2}

20 The human femur has an average cross-sectional area of 3.0 cm^2 and an unloaded length of 0.40 m.

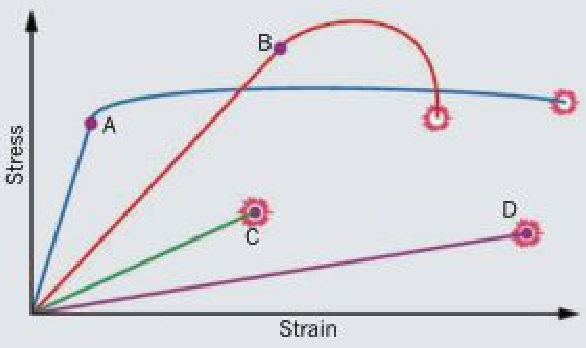
a What is the maximum compression that this bone can tolerate while behaving elastically?

b Are bones more likely to break while under tension or compression? Discuss.

c The bones of elderly people are said to be more brittle than those of younger people. Suggest some reasons why this might be so in the light of your knowledge of materials.

21 It is estimated that in sporting activities, the largest compression that a femur of length 40 cm will encounter will be about 0.30 mm. Calculate the compressive stress on the bone for this compression of the femur.

22 Use the stress–strain graph to answer the following questions.



a Which one or more of the materials is brittle?

b Which material is the stiffest?

c Which material is the least stiff?

d Which material is the strongest?

e Which material is the most ductile?

23 Explain the difference between elastic deformation and plastic deformation in terms of energy considerations.

The following data for steel apply to questions 24 and 25.

Young's modulus = 2.00×10^{11} N m^{-2}
Elastic limit = 4.20×10^8 N m^{-2}
Tensile strength = 8.20×10^8 N m^{-2}

24 A cylindrical steel rod of length 50.0 cm and radius 2.00 mm is subjected to a tensile force of 1.20 kN.

a What is the tensile stress in the rod at this tension?

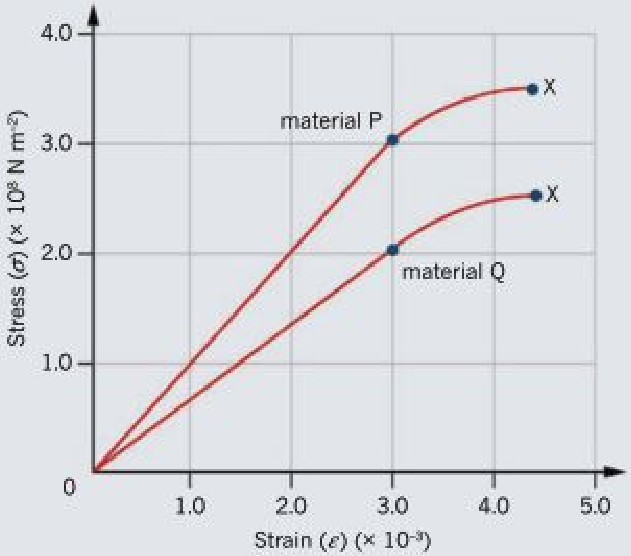
b How much strain energy per cubic metre is stored in the rod at this tension?

c What is the total amount of strain energy stored in the rod under this tension?

25 From the data provided, is it possible to obtain an accurate value for the maximum strain energy required to cause this rod to fail? Explain your answer.

The following information applies to questions 26–29.

The stress–strain graphs for two materials P and Q are shown. The points of fracture are indicated by an X.



26 Both materials have experienced a strain of 0.30%. Calculate the ratio of the strain energy per unit volume of P to the strain energy per unit volume of Q.

27 Calculate the minimum energy per unit volume that will cause failure in:

a material P

b material Q.

28 Which is the tougher material, P or Q? Justify your answer.

29 With reference to the graph, which material is:

a stiffer?

b stronger?

Justify your answers.

CHAPTER

15 Energy from nuclear power

Scientists are near universally agreed that, since the start of the industrial age, human activity has led to significant increases in the concentration of carbon dioxide in the Earth's atmosphere.

While existing and new forms of renewable energy, such as solar and wind, will play an important role in achieving carbon-free energy production, alternatives are needed to generate power which are unaffected by climatic conditions.

This chapter explores how energy is produced from nuclear fuel, how that energy is put into use for power production, and considers the viability of the alternatives both in light of the feasibility of production and the potential environmental effects.

Key knowledge

By the end of this chapter on energy from nuclear power, you will be able to:

- explain nuclear fusion reactions of proton–proton and deuterium–tritium with reference to:
  - reactants, products and energy production
  - availability of the reactants
  - energy production compared with the mass of fuel
- explain nuclear fission reactions of ^{235}U and ^{239}Pu with reference to:
  - fission initiation by slow and fast neutrons respectively
  - products of fission including typical unstable fission fragments and energy
  - radiation produced by unstable fission fragments
- describe neutron absorption in ^{238}U, including the formation of ^{239}Pu
- explain fission chain reactions including:
  - the effect of mass and shape on criticality
  - neutron absorption and moderation
- compare nuclear fission and fusion with reference to:
  - energy released per nucleon and the percentage of the mass that is transformed into energy
  - availability of the reactants
  - limitations as a source of energy for electricity production
  - environmental impact
- analyse fission and fusion with reference to their viabilities as energy sources
- describe the energy transfers and transformations in the systems that convert nuclear energy into thermal energy for subsequent power generation
- explain the risks and benefits for society of using nuclear energy as a power source.

15.1 Energy from the nucleus

The generation of power from **nuclear power stations** is in reality very similar to the generation of power from coal or even water. The power source is the essential difference in each case.

FIGURE 15.1.1 Chapelcross Nuclear Power Station, Annan, Dumfries, Scotland.

Power can be derived by the direct use of nuclear fuel using two basic methods—fission and fusion. Fission is already in widespread use in many countries around the world (Figure 15.1.1). A nuclear fission power plant uses the heat generated by a nuclear fission process to drive a steam turbine that generates usable electricity. Fusion is the method by which the Sun produces its power. In this section, the processes by which fission and fusion produce energy are explained.

PHYSICSFILE

Geothermal energy

While this section focuses on the direct production of nuclear energy from fission and fusion, they aren't the only sources of energy from nuclear processes. A less obvious source of energy from nuclear processes is the heat from naturally occurring radioactive decay within Earth itself. It is a significant source of the heat in the interior of the Earth. Radioactive isotopes of uranium, thorium and potassium are the primary contributors. The heat from radioactive decay is the primary source of heat on which geothermal energy relies. Iceland's unique geological features enable it to harness geothermal energy. Geothermal power plants currently account for over 25% of the nation's electricity production.

FIGURE 15.1.2 A geothermal power station at Svartsengi, Iceland. In the foreground, bathers enjoy the benefits of the 'Blue Lagoon'.

ENERGY FROM FISSION

In Chapter 7 'Energy from the atom', the process by which a heavy nucleus splits on impact with another particle was described. In a fission reactor there must be processes that initiate fission and then maintain and control the reaction.

Enrico Fermi realised that neutrons would be the most effective means of initiating nuclear reactions. They have no net charge and so are not repelled by a positively charged nucleus or by an alpha particle. This means that the chance of a neutron reaching the nucleus of an unstable isotope is much greater than the chance of a charged particle reaching the nucleus, as shown in Figure 15.1.3.

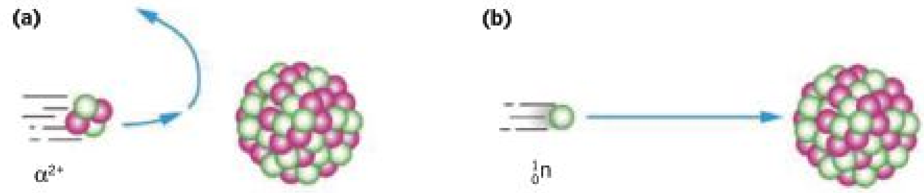


FIGURE 15.1.3 (a) Alpha particles are repelled so strongly by large atomic nuclei that collisions are not possible. (b) Neutrons, having no charge, are capable of colliding directly with the nucleus of an atom.

The products of fission include two or three neutrons, a little more on average from the fission of Pu-239 than from U-235. So, if fission could be initiated, then the reaction could become self-sustaining as the neutrons from one fission process initiate further processes.

Slow and fast neutrons

There are problems that need to be overcome to initiate the fission process. In a **nuclear fission reactor**, a neutron source is inserted into the **core** of the reactor in regularly spaced positions to initiate the fission reaction. Neutrons emitted during fission are moving very fast. They typically have energies of between 0.1 and 3 MeV.

This high kinetic energy is useful in that it allows a fast neutron to split more nuclei once they get captured. This happens alongside the splitting of other by-products of the fission process such as some **actinides** that build up in reactors as radioactive waste.

Fast neutrons are able to fission other heavy nuclei present in the reactors, including U-238, the most abundant isotope in naturally occurring uranium. So called **fast reactors** or fast-breeder reactors take advantage of this (see Figure 15.1.4). That is, they make use of uranium-238 which is abundant, but is effectively non-fissile.

FIGURE 15.1.4 The Superphénix commercial-size reactor is the largest fast-breeder reactor in the world. It is located in Creys-Malville in France.

These fast-breeder reactors work by first changing U-238 into another substance. The absorption of at least one neutron per fission by U-238 nuclei transforms it into fissile plutonium, Pu-239. This process is known as 'breeding' since it produces more high-fissile Pu-239 than the uranium it uses as fuel. The process is also useful in burning long-half-life actinides that are an unwanted constituent of radioactive waste.

The process of neutron absorption by U-238 to form Pu-239 takes several stages, as seen in Figure 15.1.5. A U-238 nucleus capturing a neutron forms U-239. This is a highly unstable isotope with a half-life of 23 minutes which transforms via beta emission to Np-239 (neptunium-239). Np-239 also decays via beta emission, with a half-life of 3 days on average, to become the fissile Pu-239 (plutonium-239). Pu-239 is a good fission fuel, but can also be used in nuclear weapons.

This process can be shown as follows:

$$^{238}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{239}_{92}\text{U} + \gamma$$

$$^{239}_{92}\text{U} \xrightarrow{\beta^-} ^{239}_{93}\text{Np} \xrightarrow{\beta^-} ^{239}_{94}\text{Pu}$$

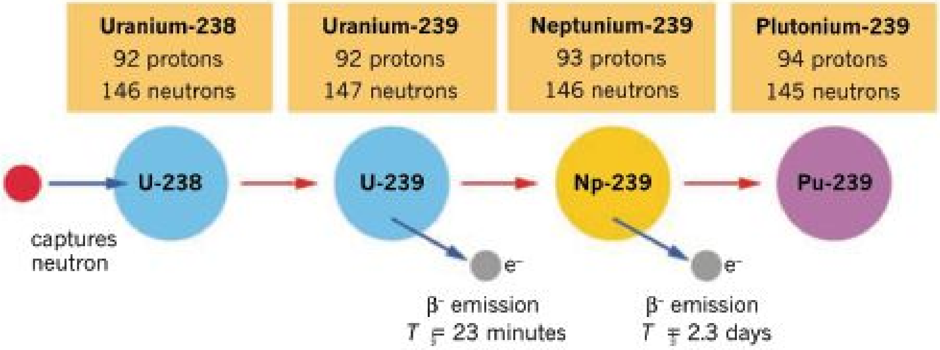


FIGURE 15.1.5 Stages in which uranium-238 is transformed into plutonium-239.

Plutonium requires fast-moving neutrons to bring about nuclear fission, so it is ideal in fast-breeder reactors where the emitted neutrons travel at very high speeds.

Use of moderators to slow down neutrons

Unlike plutonium, the probability that a U-235 nucleus will absorb neutrons and undergo fission is highest when the neutrons are moving slowly. In a standard thermal reactor, ensuring enough slow neutrons exist within a reactor is vital for the fission process to continue. A substance referred to as a **moderator** is used to slow the neutrons. The most effective moderators will have atoms with a mass as close as possible to that of a neutron—a neutron hitting a larger mass would bounce off with almost as much energy as when it hits.

Graphite and **heavy water** both make effective moderators for this reason. In a nuclear reactor, once neutrons have been slowed down by a number of collisions with atomic nuclei, they are referred to as *slow* neutrons. The term is a little misleading as even then, the neutrons may be moving at 2000 m s^{-1}. Their speed is slow compared to the speed of light, and the kinetic energy is considerably less than at the speeds at which the neutrons are emitted. Slow neutrons have a kinetic energy of around 0.025 eV.

Natural uranium only contains about 0.7% of fissile U-235. A reactor using this mix as a fuel must use heavy water or graphite as a moderator to capture very few neutrons so that fission can continue. By enriching the fuel to around 3–4% U-235, ordinary water (either boiling or under pressure) can be used as a moderator. The moderator (water) both slows the neutrons and cools the reactor core. The U-238 in the uranium fuel will be converted over time to fissile Pu-239 via neutron absorption as described above.

Chain reactions and critical mass

As mentioned above, a fission reactor will have a neutron source to initiate fission and neutrons in the reactor need to be slowed by a moderator. This is so that sufficiently large numbers of neutrons can be absorbed by uranium nuclei for the fission reaction to continue. There are other factors that determine whether fission is initiated and then continues; these will now be considered.

In designing a fission device, whether it is a nuclear weapon or a nuclear reactor, the intent is to produce a self-sustaining **chain reaction**. In a nuclear weapon, the process is uncontrolled–the whole idea is for it to go off with a 'bang'. In a reactor, **control rods** are used to control the reaction by absorbing neutrons to reduce the number available for fission reactions. By increasing the surface area of the control rods and by inserting or removing the rods, operators of the reactor can increase, decrease or even fully stop the fission process (see Figure 15.1.6).

Two or three neutrons are released in each fission of a uranium nucleus. Each of these can potentially be absorbed by another nearby nucleus causing two or three more nuclei to fission. The number of nuclei undergoing fission at least doubles with each generation, and within a small fraction of a second an enormous number of nuclei will have undergone fission. Only a very small amount of energy (of the order of 10^{-13} J) is released by each fission reaction, but in this chain reaction there are so many reactions occurring in such a short time that massive amounts of energy can be released. This process is shown in Figure 15.1.7.

FIGURE 15.1.6 Inside a nuclear power station. The coloured squares provide access to the core that contains rods of uranium. Control rods are inserted into the reactor via the coloured squares to control the amount of heat generated.

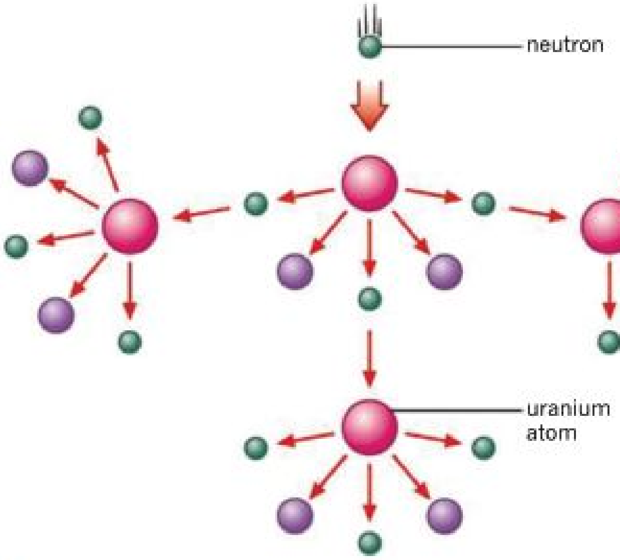


FIGURE 15.1.7 A nuclear fission chain reaction: a neutron (green) is absorbed by the U-235 nucleus (pink) causing it to split into two smaller nuclei (purple) and releasing, in this case, three neutrons. These are in turn absorbed, causing more nuclei to split. The rapid increase in free neutrons created by each successive splitting creates a chain reaction, releasing enormous amounts of energy.

The amount of nuclear fuel required to initiate continuous fission, or 'go critical' through a chain reaction, depends on a number of factors: the purity of the fissile material, its shape and its mass.

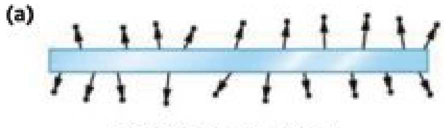
Purity of the nuclear material

A chain reaction cannot be established in a sample of nuclear material in which the concentration of U-235 or Pu-239 is too low. Neutrons have only a small chance of being absorbed by fissile nuclei and causing further fission reactions. The greater the proportion of fissile material, the greater the chance that a fission chain reaction will continue.

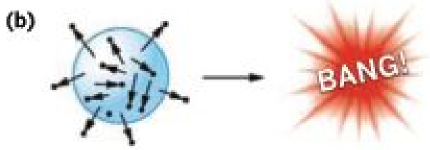
Shape of the nuclear material

The shape of the nuclear fuel is an important factor in determining whether a chain reaction will occur. A 20 kg sample of enriched U-235 in the shape of a sphere can potentially spontaneously explode, whereas the same mass of enriched U-235 flattened into a sheet will not (see Figure 15.1.8).

The flat piece has a very large surface area, so a large number of the neutrons emitted are able to escape from the uranium into the air. These neutrons won't be available to cause further fission reactions and so the chain reaction dies out. A spherical piece of uranium has a substantially smaller surface area so a greater proportion of neutrons remain in the uranium to sustain the chain reaction.



20 kg ^{235}U: subcritical



20 kg ^{235}U: supercritical!

FIGURE 15.1.8 (a) The large surface area of the flat piece of uranium-235 enables a large proportion of neutrons to escape into the air, causing the chain reaction to die out. (b) In the spherical piece of uranium-235, a sufficient proportion of neutrons remain inside the material to maintain the chain reaction, leading to an explosion.

Mass of the nuclear material

The ability of a fissile material to maintain a chain reaction depends on its mass. The smallest mass that is just capable of exploding is called the **critical mass**.

In a fission reaction when a high proportion of neutrons escape or are absorbed by atoms other than U-235, the chain reaction is not sustained. This is a subcritical situation (Figure 15.1.9(a)), and has a **subcritical mass**. When there are just enough neutrons available to keep the chain reaction going at a steady rate, the situation is described as critical (Figure 15.1.9(b)). When a high proportion of released neutrons are available to continue the fission process so that the chain reaction grows, it is described as **supercritical** (Figure 15.1.9(c)). Nuclear weapons make use of supercritical reactions.

The aim of a nuclear reactor is to maintain nuclear materials at critical mass.

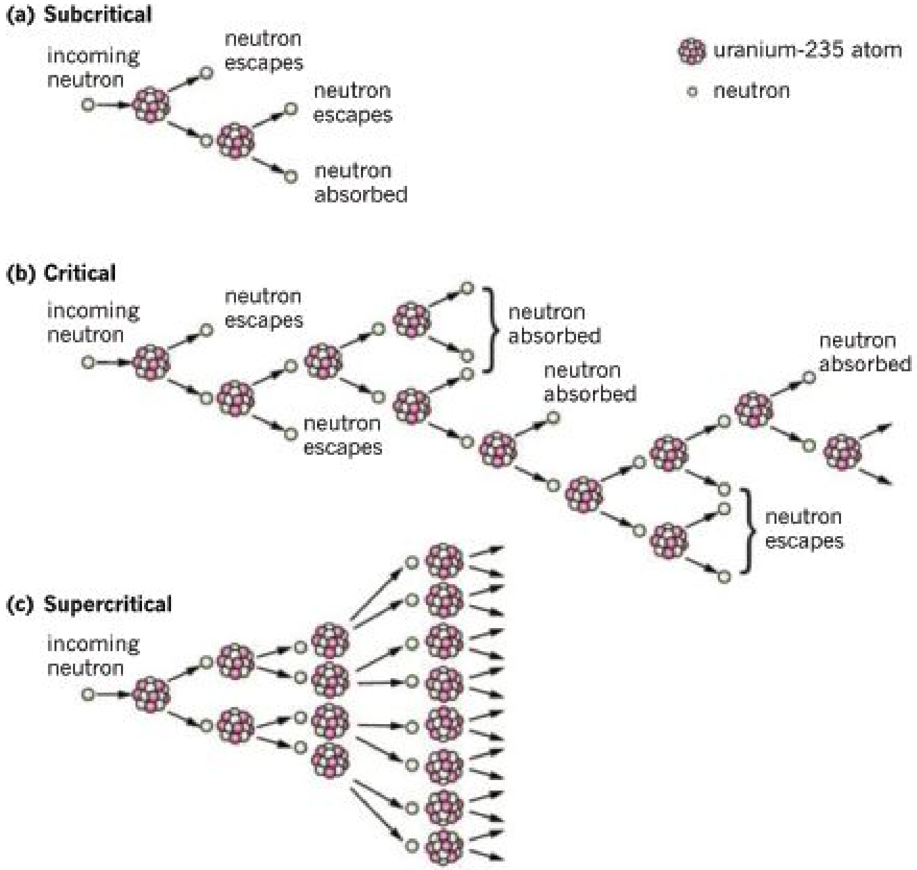


FIGURE 15.1.9 Fission reactions fall into three categories: subcritical, critical and supercritical.

The actual mass of a fissile material that is needed to reach critical mass is not an absolute value. It depends upon the factors already discussed (shape and purity) along with the sample's density and the temperature of the surroundings. Hot fuel is always less reactive than cold fuel as it will be less dense due to thermal expansion and will be of higher energy overall due to thermal energy. Neutrons will be released with higher average energies, reducing the probability that they will be absorbed.

Combining both critical mass and shape determines the critical size of fissile material required for the core of a reactor. The critical size must include at least enough fissionable material for it to reach critical mass. Less than this size and a chain reaction cannot be sustained. The smallest critical size will be achieved with a sphere, as it has the smallest surface area for a given volume of material. The critical mass of a sphere of pure U-235 is 52 kg and can be contained in a sphere 17 cm in diameter. For Pu-239 it is 10 kg and a sphere of just under 10 cm in diameter. For lower-grade fuel, the mass required rapidly rises. 400 kg of 20% pure U-235 would be needed to reach critical mass and for 15% pure U-235 the amount increases to more than 600 kg.

PHYSICS IN ACTION

Nuclear weapons

A nuclear weapon must contain enough fissile fuel for an uncontrolled chain reaction to be sustained (a supercritical situation) but the mass must be kept subcritical until detonation is required. In early weapons, this was achieved by firing one mass of fuel down a gun barrel (see Figure 15.1.10) into another piece of the fuel.

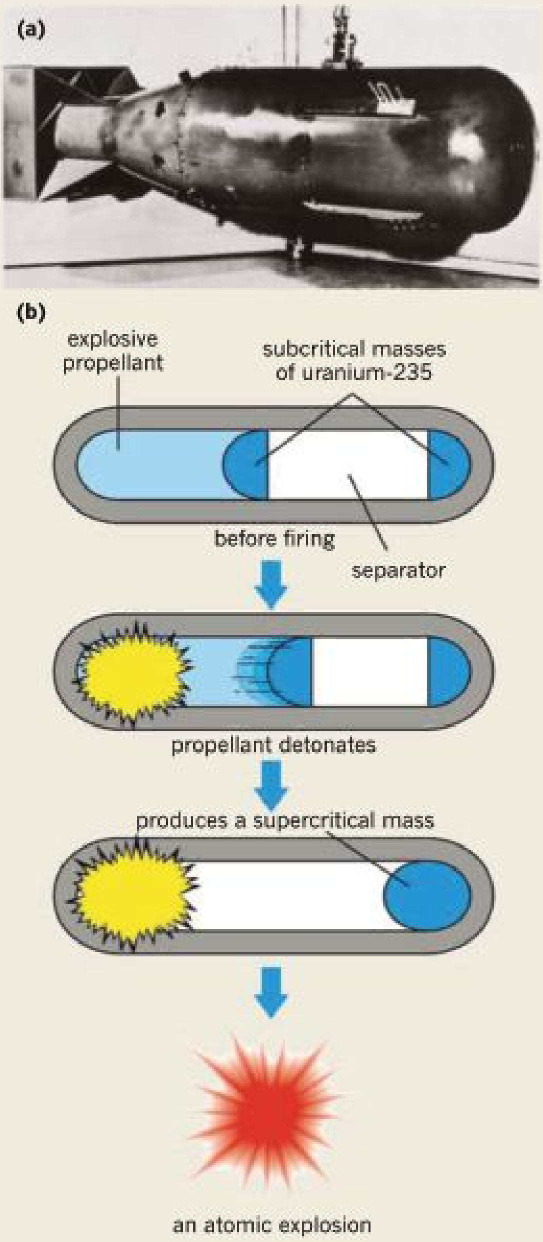


FIGURE 15.1.10 A gun-barrel type nuclear weapon. (a) A replica of Little Boy, the uranium-235 bomb that was dropped on Hiroshima on 6 August 1945. (b) A simple fission bomb contains two subcritical pieces of uranium-235. The two pieces are kept separate until they are forced together, creating a supercritical mass and leading to an explosion within one-millionth of a second.

In plutonium-based weapons, it is possible that the smaller mass could go critical before the required time due to small traces of Pu-240, which has a tendency to spontaneously fission (Figure 15.1.11). In these weapons, a charge of conventional explosive is 'shaped' around a subcritical mass of plutonium. When the shaped charge is detonated it causes the plutonium to implode, increasing the density and causing the plutonium to go critical.

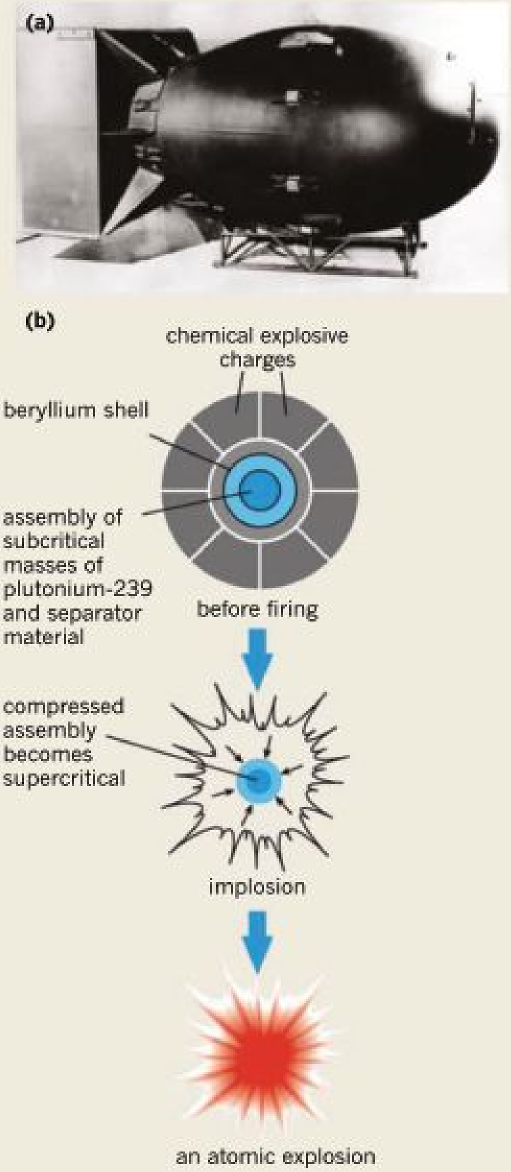


FIGURE 15.1.11 (a) A replica of Fat Man, the plutonium bomb that was dropped on Nagasaki on 9 August 1945. (b) Fat Man was a spherical fission bomb. In this type of bomb, an implosion forces a large number of subcritical pieces of plutonium-239 into one supercritical mass, producing a nuclear explosion. The beryllium shell reduces the loss of neutrons from the core.

Fission products

Once a slow neutron is absorbed by a U-235 nucleus, the energy from that neutron is distributed among all of the now 236 neutrons and protons, creating an unstable nucleus (see Figure 15.1.12). The nucleus is now likely to fission into two relatively even fragments at a mass ratio usually of around 7:10, along with the 2 or 3 neutrons that keep the fission process going.

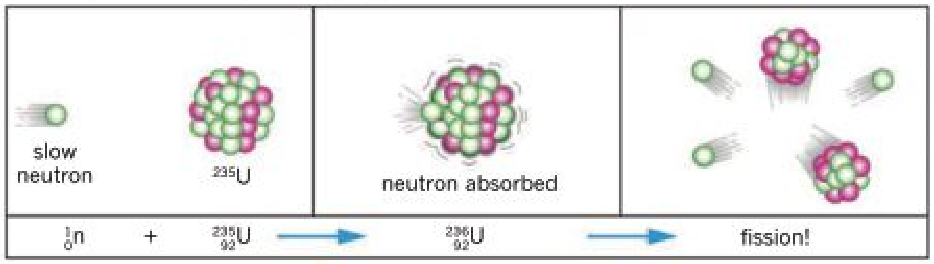


FIGURE 15.1.12 A slow neutron is absorbed by a uranium-235 nucleus, converting it into uranium-236, which is highly unstable. This nucleus then undergoes nuclear fission.

Exactly what products are produced will vary with the several hundred possible combinations likely to be created in any specific fission. While the products can be predicted by statistical probability, they cannot be definitely determined. However, the total number of neutrons and the total energy must be conserved. Many of the products will be unstable isotopes of the particular elements produced and so will decay to form more stable isotopes.

Some typical examples demonstrating this process include:

$$\text{U-235} + \text{n} \rightarrow \text{Ba-144} + \text{Kr-90} + 2\text{n} + 200\text{ MeV}$$

$$\text{U-235} + \text{n} \rightarrow \text{Ba-143} + \text{Kr-90} + 3\text{n} + 170\text{ MeV}$$

$$\text{U-235} + \text{n} \rightarrow \text{Zr-94} + \text{Te-139} + 3\text{n} + 197\text{ MeV}$$

In each case, the total numbers of neutrons and protons are conserved—the small mass defect will be equivalent to the energy released.

The barium and krypton isotopes in the examples above are unstable isotopes of each of these elements. Considering the first example shown above, Ba-144 will decay over several steps to form a stable isotope of neodymium; Kr-90 decays to form a stable isotope of zirconium. In each case the decay will involve the emission of several beta particles (electrons) and it is this beta decay, along with the associated gamma emissions, that makes the material highly radioactive.

The decay of Ba-144 to neodymium can be shown as:

$$\text{Ba-144} \xrightarrow{\beta} \text{La-144} \xrightarrow{\beta} \text{Ce-144} \xrightarrow{\beta} \text{Pr-144} \xrightarrow{\beta} \text{Nd-144}$$

Kr-90 decays to a stable isotope of zirconium via the following process:

$$\text{Kr-90} \xrightarrow{\beta} \text{Rb-90} \xrightarrow{\beta} \text{Sr-90} \xrightarrow{\beta} \text{Y-90} \xrightarrow{\beta} \text{Zr-90}$$

Similarly, other potential products and the decay of those products can be described.

As has already been noted, most of the uranium in a nuclear reactor is U-238. It will capture neutrons to become U-239 and then, through the process described in the previous section, will become Pu-239. Over a period of about three years, enough Pu-239 will be produced that almost one-third of the energy output will come from the fission of this Pu-239. One such fission of Pu-239 is as follows:

$$\text{Pu-239} + \text{n} \rightarrow \text{Ce-148} + \text{Kr-89} + 3\text{n} + 210\text{ MeV}$$

There are many hundreds of possible products from the fission of Pu-239.

Used fuel includes isotopes of plutonium, curium, neptunium and americium. Each of these are alpha-emitters with long half-lives comparable to uranium isotopes. This is why used fuel needs storage far beyond the few thousand years that would be necessary solely for the decay of fission products (Figure 15.1.13).

FIGURE 15.1.13 The underground storage and testing of containers of radioactive wastes in New Mexico, USA.

ENERGY FROM FUSION

The total mass of every stable atomic nucleus is less than the total mass of the individual protons and neutrons. This means that if, say, two protons and two neutrons were 'fused' together to form a helium nucleus as shown in Figure 15.1.14, there would be a loss of mass. You saw in Chapter 7 'Energy from the atom' that this mass deficit would lead to the release of a very large amount of energy. The potential for large amounts of clean energy with little or no radioactive by-products has led to intense research into the feasibility of the widespread commercial production of energy from nuclear fusion.

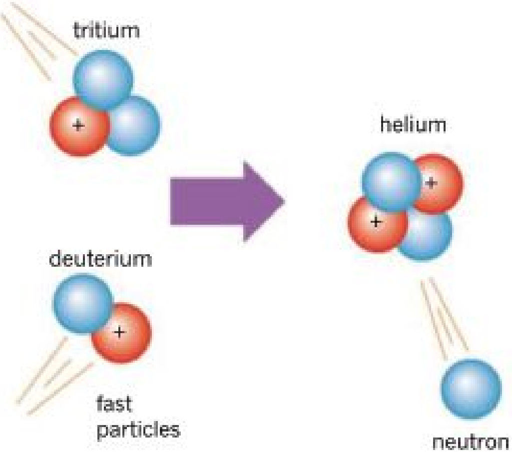


FIGURE 15.1.14 A deuterium–tritium reaction that could be used to provide fusion power. A deuterium (hydrogen-2) nucleus fuses with a tritium (hydrogen-3) nucleus to form a helium-4 nucleus and emits a neutron.

Fusion reactions

One of the simplest fusion reactions involves the production of an isotope of hydrogen, deuterium or H-2, from a neutron and a proton as follows:

$$^{1}_{1}\mathrm{H} + \mathrm{n} \rightarrow {}^{2}_{1}\mathrm{H} + \gamma$$

The mass defect after the reaction is equivalent to an energy release of 2.22 MeV per deuterium nucleus produced, which is carried off by the nucleus and by the gamma radiation.

The principle process responsible for the energy output from the Sun is believed to be from a proton–proton cycle in which four protons are combined to form one helium nucleus along with two positrons (e^+), two neutrinos (ν), two gamma rays and 26.7 MeV of energy.

The sequence of fusion reactions including the energy released is as follows:

$$^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{2}_{1}\mathrm{H} + e^{+} + \nu + 0.42\ \mathrm{MeV}$$

$$^{2}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + \gamma + 5.49\ \mathrm{MeV}$$

$$^{3}_{2}\mathrm{He} + {}^{3}_{2}\mathrm{He} \rightarrow {}^{4}_{2}\mathrm{He} + {}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} + 12.86\ \mathrm{MeV}$$

It takes two of each of the first two steps to produce the two $^{3}_{2}\mathrm{He}$ for the third reaction, so the energy released is not a straight addition of the energy for each individual step. The net effect of these combinations of steps is:

$$4\,{}^{1}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He} + 2e^{+} + 2\nu + 2\gamma + 24.7\ \mathrm{MeV}$$

Each of the two positrons (e^+) quickly annihilates with an electron (e^-) to release an additional 2 × 1.02 MeV, accounting for the additional energy making up the total energy released of 26.7 MeV.

The first step in this process, the formation of deuterium, has a very low probability, which limits the rate at which the Sun can produce energy. This reaction can be seen in Figure 15.1.16.

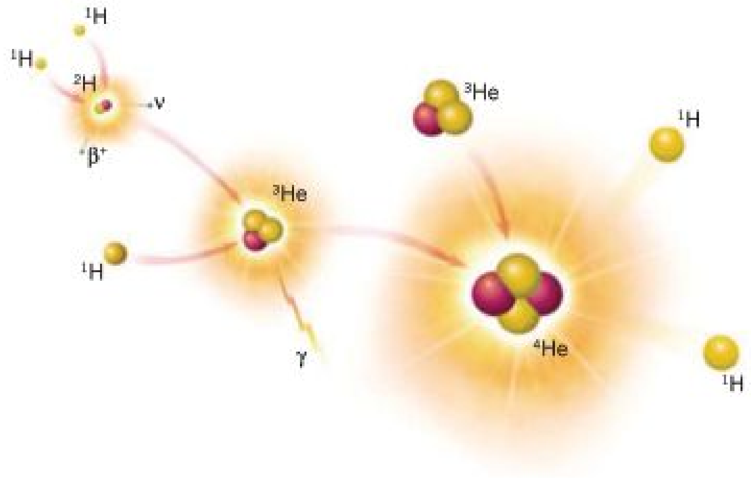


FIG 15.1.16 The proton–proton reaction.

Research into fusion reactions suggests that this simple proton–proton reaction could use the plentiful supply of the H-1 (protium) isotope of hydrogen that makes up the bulk of naturally occurring hydrogen. 99.9844% of the hydrogen occurring in nature is protium. However, the very low probability of such a fusion reaction happening would make it extremely unlikely to succeed.

PHYSICSFILE

Hydrogen isotopes

There are three isotopes of hydrogen: hydrogen-1, hydrogen-2, and hydrogen-3. Each isotope has a particular name:

H-1: protium

H-2: deuterium

H-3: tritium

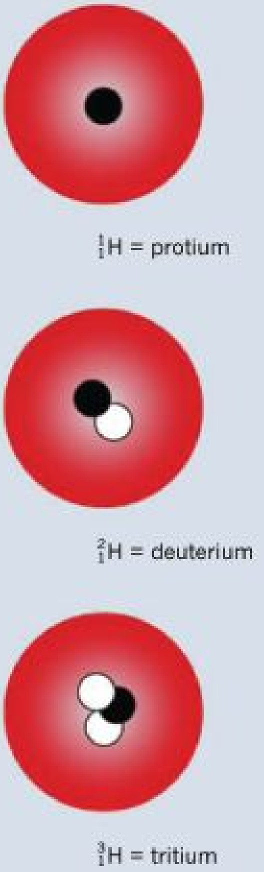


FIGURE 15.1.15 The three isotopes of hydrogen.

The most likely reaction to succeed in a practical fusion reactor is the deuterium–tritium (H-2 and H-3 isotopes of hydrogen, respectively) reaction:

$$^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{1}H + ^{1}_{1}H + 4.03\ \text{MeV}$$

$$^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{2}He + n + 3.27\ \text{MeV}$$

$$^{2}_{1}H + ^{3}_{1}H \rightarrow ^{4}_{2}He + n + 17.59\ \text{MeV}$$

Comparing this energy output to the fission of U-235, potentially 3.38×10^{14} J of energy could be produced for each kilogram of deuterium–tritium fuel compared with 8.80×10^{13} J per kilogram of uranium—almost four times the energy yield per kilogram. Also, deuterium, while not as abundant as protium, still makes up 0.0156 % of the hydrogen found in nature. That might not sound like a lot but it means that for every 60 litres of water in the world's oceans, 1 g of deuterium can be extracted—an extremely large fuel source when the total volume of water in the world's oceans is considered.

The technology to produce a deuterium–tritium reactor was proven in 1997 in a small reactor called the Joint European Torus (JET) or tokamak. The name refers to the shape of the magnetic field or torus used to confine the reaction (Figure 15.1.17). However, there are considerable technical hurdles to overcome before the process could be considered for wide-scale deployment.

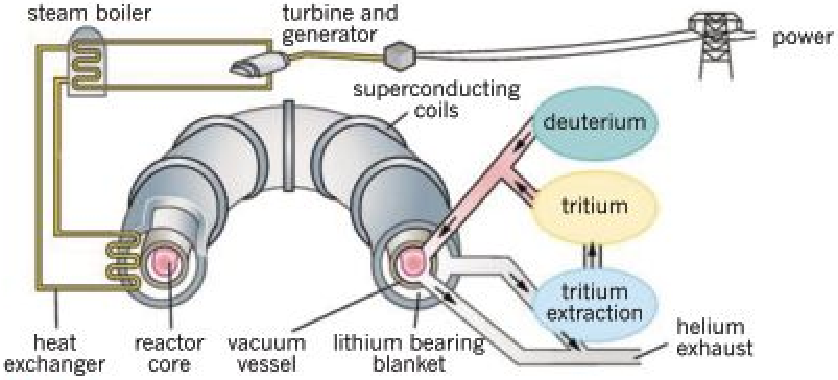


FIGURE 15.1.17 This design of a tokamak fusion reactor shows the doughnut shape of the core where the plasma of fusion fuel and products should be situated. The fusion reaction is held in place by a powerful magnetic field, and lithium and water are used to extract the heat energy from the core. It is intended that this energy could then be used to generate enormous amounts of electricity.

Once hydrogen nuclei get very close together, attractive nuclear forces pull the nuclei together. But to initially get that close, the repulsive forces of the like-charged protons in each nucleus must first be overcome. For that, the nuclei must have very large kinetic energies such as those coming from the very high temperatures found in the core of the Sun. In fact, the required kinetic energy of approximately 0.22 MeV would require a temperature in the order of 4×10^{7} (40 million) kelvin.

In addition, while deuterium is relatively common, tritium does not exist in nature as it is unstable with a 10-year half-life and would have to either be generated from a deuterium–deuterium cycle (the first step in the deuterium–tritium cycle above), or bred from lithium-6 by neutron bombardment:

$$^{6}_{3}Li + ^{1}_{0}n \rightarrow ^{4}_{2}He + ^{3}_{1}H + 4.8\ \text{MeV}$$

This can be achieved with slow neutrons and would occur if lithium were used as the coolant around the reaction chamber of a fusion reactor. Only 7.4% of naturally occurring lithium is Li-6, a sizeable supply but substantially less than the supply of deuterium available. This would be the limiting resource for a fusion industry based on deuterium–tritium reactors.

15.1 Review

SUMMARY

- In a fission reactor, a neutron source is inserted into the core of the reactor in regularly spaced positions to initiate the fission reaction.
- Neutrons emitted during the fission process are referred to as fast neutrons. Fast neutrons can split more nuclei once they get captured, along with the splitting of other by-products of the fission process, such as some actinides that build up in reactors as radioactive waste.
- Fast-breeder reactors use nuclear fuel enriched in fissile material to transform U-238 into fissile P-239 and produce more P-239 than they use U-238.
- The process of neutron absorption by U-238 to form Pu-239 takes several stages and can be shown as:
  $^{238}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{239}_{92}\text{U} + \gamma$
  $^{239}_{92}\text{U} \xrightarrow{\beta^-} ^{239}_{93}\text{Np} \xrightarrow{\beta^-} ^{239}_{94}\text{Pu}$
- The probability that a U-235 nucleus will absorb neutrons is highest when the neutrons are moving slowly. In a standard reactor, ensuring enough slow neutrons exist within a reactor is vital to the fission process continuing. A moderator is used to slow the neutrons.
- Once a slow neutron is absorbed by a U-235 nucleus, the energy from that neutron is distributed among all of the now 236 neutrons and protons, creating an unstable nucleus that will fission into two relatively even fragments. Many of the products will be unstable isotopes that will subsequently decay. Many hundreds of possible combinations of fragments are possible.
- The major proportion of uranium in a nuclear reactor is U-238. It will capture neutrons to become U-239 and then Pu-239, which will also fission. After about 3 years, the fission of Pu-239 will produce almost a third of the total power output.
- The total number of neutrons and protons in the fragments are conserved—the small mass defect will be equivalent to the energy released.
- The decay of unstable fragments will involve the emission of several beta particles (electrons), and it is this beta decay along with the associated gamma emissions that make the material highly radioactive.
- In a fission reaction, at least one neutron per fission is released. The free neutrons may collide with further fissile nuclides and a chain reaction may occur. The more neutrons released per fission, the higher the probability that the chain reaction will be sustained.
- In a sample of nuclear material in which the concentration of U-235 or Pu-239 is too low, a chain reaction cannot be established.
- The ability of a fissile material to maintain a chain reaction depends on its mass. The smallest mass that is capable of exploding is called the critical mass.
- The mass of a fissile material needed to reach critical mass is not an absolute value. It depends on the material's shape, purity and density and the temperature of the surroundings.
- The critical size of a sample of fissionable material contains enough material for it to reach critical mass. The smallest critical size will be achieved with a sphere as it has the smallest surface area for a given volume of material.
- The principle process believed responsible for the energy output from the Sun is believed to be from a proton–proton cycle in which four protons are combined to form one helium nucleus along with two positrons, two neutrinos, two gamma rays and 26.7 MeV of energy.
  $4^{1}_{1}\text{H} \rightarrow ^{4}_{2}\text{He} + 2\text{e}^+ + 2\nu + 2\gamma + 24.7\ \text{MeV}$
- Each of the two positrons (e^+) annihilates with an electron (e^-) to release an additional 2 × 1.02 MeV, accounting for the additional energy making up the total energy released of 26.7 MeV.
- The most likely reaction to succeed in a practical fusion reactor here on Earth is the deuterium–tritium reaction:
  $^{2}_{1}\text{H} + ^{3}_{1}\text{H} \rightarrow ^{4}_{2}\text{He} + \text{n} + 17.59\ \text{MeV}$
- The energy yield per kilogram of fuel for a deuterium–tritium fusion reaction is almost four times that for the fission of U-235.
- The supply of deuterium as a fuel is almost unlimited; however, factors that limit the successful deployment of fusion reactors include the lack of naturally occurring tritium, requiring 'breeding' from less-abundant lithium-6, and the very high temperatures required to initiate fusion, in the order of 40 million kelvin.

KEY QUESTIONS

1 Pu-239 is an important fissile fuel in a nuclear reactor. It is present in a nuclear reactor as a result of which of the following?
   - **A** naturally occurring deposits mixed with uranium
   - **B** fission of U-235
   - **C** neutron absorption by U-235
   - **D** neutron absorption by U-238

2 Fission reactors require a moderator, often in the form of graphite or heavy water. What is the function of a moderator?
   - **A** to absorb excess neutrons emitted via fission
   - **B** to absorb radioactive waste emitted via fission
   - **C** to decrease the energy of neutrons emitted via fission
   - **D** to increase the energy of neutrons emitted via fission

3 A fast-breeder reactor is a simpler design than a thermal nuclear reactor.
   - **a** What fuel is used in fast breeder reactors?
   - **b** Why is it called a 'fast' reactor?
   - **c** Why is it called a 'breeder' reactor?

4 Fast neutron absorption by U-238 to form Pu-239 is the first step in a multistep process to form Pu-239. Complete each step with the type of decay or absorption required for each process.
   - **a** U-238 to U-239
   - **b** U-239 to Np-239
   - **c** Np-239 to Pu-239

5 During fission, U-235 releases an average of 2.47 neutrons while Pu-239 releases an average of 2.89 neutrons.
   - **a** What impact does the number of neutrons produced per fission have on the production of energy?
   - **b** Which nuclide, U-235 or Pu-239, would establish the stronger chain reaction if used as nuclear fuel?

6 The critical mass of U-235 is about 50 kg. Explain why a similar mass of U-235 flattened into a thin sheet would not sustain a fission reaction.

7 Complete the sentence about fission by choosing the correct responses from the ones in brackets.
The fission of U-235 forms **[one/two/three]** relatively even fragments along with 2 or 3 **[protons/neutrons/electrons]** that keep the fission process going. What products are produced will **[disappear/remain constant/vary]**. However, the total number of neutrons and total energy must be **[changed/conserved]**.

8 The energy output of the Sun is believed to be mainly from a proton–proton cycle. A proton–proton fusion reaction in the Sun can be summarised as:
$4{}^{1}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + 2e^{+} + 2\nu + 2\gamma + 24.7\text{ MeV}$
The total energy produced in this process is noted as 26.7 MeV. Where does the additional energy outside of that from the fusion reaction come from?

9 What is the fusion reaction most likely to succeed on a practical level on Earth?

10 Describe two limiting factors in the deployment of deuterium–tritium fusion reactors.

15.2 Nuclear energy as a power source

The Nuclear Energy Institute, a body advocating for the peaceful and beneficial use of nuclear energy in the United States, reported that as of July 2015 there were 438 nuclear power stations in operation around the world in a total of 30 countries. 10.9% of the world's total electricity production came from nuclear power and more than 13 countries relied on nuclear power for more than one-quarter of their energy needs, with France topping the list. Over three-quarters of France's electrical power comes from nuclear power. Australia is yet to build a single nuclear power station but does have a research facility at Lucas Heights, Sydney.

A further 67 nuclear power stations were under construction in 15 countries. This growth comes despite the nuclear disasters at Chernobyl, Three Mile Island and, more recently, Fukushima, Japan, in 2011 (see Figure 15.2.1), as countries need new, low-carbon energy sources. In this section, you will examine how nuclear power is harnessed to produce electricity, and the associated risks and benefits.

FIGURE 15.2.1 The Fukushima Daiichi nuclear power station on Japan's Pacific coast was badly damaged in the March 2011 tsunami.

NUCLEAR POWER GENERATION

At present, only small experimental reactors are attempting to prove the viability of nuclear fusion as a power source for electrical generators. Every nuclear power station currently in operation and under construction uses nuclear fission. Regardless of the process, the fundamental difference between a nuclear power station and a coal-fired power station is the fuel source. A thermal nuclear power station replaces coal with nuclear fission as the heat source to generate steam to turn a turbine.

Thermal nuclear power stations

While there are differences in design and implementation, all fission-powered thermal nuclear power stations have several key common features (see Figure 15.2.2).

FIGURE 15.2.2 Nuclear power plant at night in Grohnde, Germany. On the left are the cooling towers with water vapour; on the right is the reactor.

The most common type of nuclear power station currently in operation is the pressurised water reactor, as seen in Figure 15.2.3 on page 545. By using uranium fuel enriched to contain 3–4% U-235, ordinary water under pressure can be used as the moderator to slow neutrons from fission reactions. A pressurised water reactor does not need graphite or heavy water as a moderator. The primary difference between this and a coal-fired generator is in the way the heat is produced—a nuclear reactor uses the fission process and a coal generator burns coal.

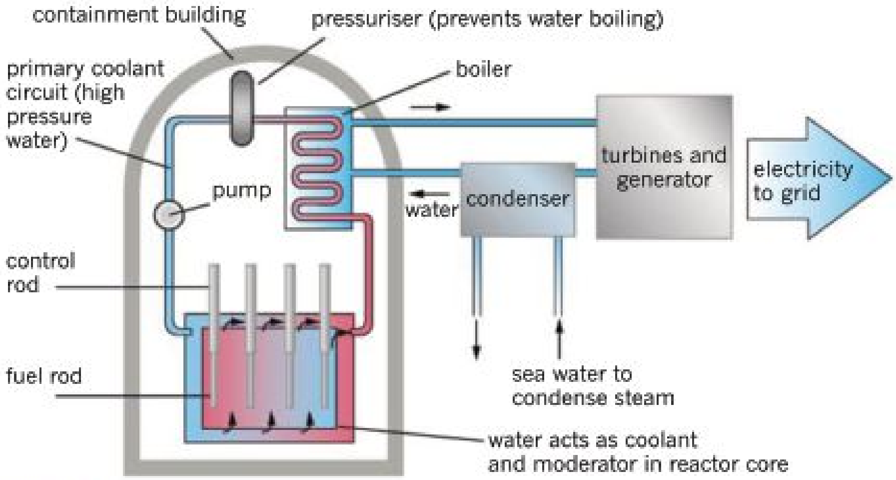


FIGURE 15.2.3 A schematic diagram of a pressurised water reactor.

Surrounding the reactor is a containment structure. This is intended to contain any radioactivity that may be released from the reactor. At the centre of the reactor is the core containing the uranium fuel in **fuel rods**, enriched in U-235. The **coolant** keeps the core from overheating.

When a U-235 nucleus undergoes fission, fast neutrons are released. Pressurised water acts as a moderator, slowing the neutrons and increasing the probability that the neutrons will be absorbed by another U-235 nucleus. Fission and neutron emission continue in a chain reaction, generating large amounts of heat energy in the process and heating the pressurised water. Being under high pressure, the water will remain a liquid.

The rate of the chain reaction can be controlled by raising or lowering control rods into the core. Control rods contain boron-10. Boron-10 has a high neutron absorption capability. By changing the position and number of control rods, the number of neutrons available for fission can be controlled, as shown in Figure 15.2.4. This allows the operators of the reactor to reduce or shut down the reactor at times of low demand or for maintenance.

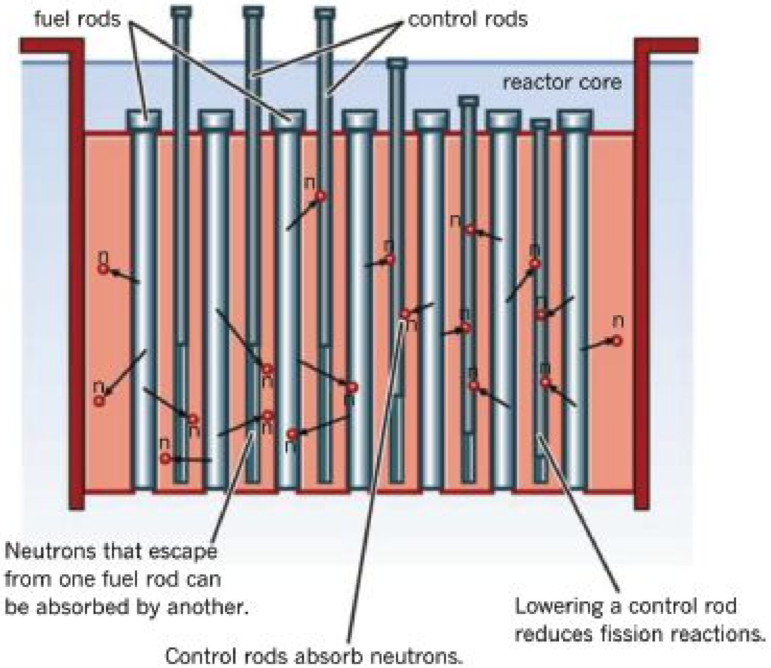


FIGURE 15.2.4 Diagram showing control rods positioned between fuel rods inside a nuclear reactor core.

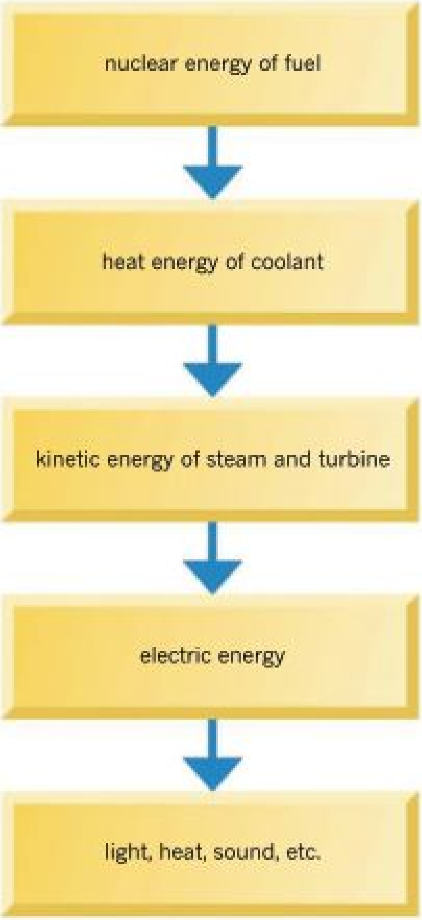


FIGURE 15.2.5 Flow chart of the energy transformations involved as a fission reactor is used to produce electricity.

Up until this stage, the process is unique to thermal nuclear reactors. From this point the process is little different to a coal-fired power station. Outside the core, the hot, pressurised water is pumped through a **heat exchanger**, converting cold water in the external pipes to steam, driving a standard electrical turbine and generating electrical power to the grid.

There are several energy transformations through this process that are highlighted in the flow chart in Figure 15.2.5. Each energy transformation can be linked to part of the process illustrated in Figure 15.2.3 on page 545.

FISSION VERSUS FUSION

It's been noted already that despite many years of research, fusion reactors are yet to reach the stage of practical implementation. So far, fusion has only been initiated on a very small scale for brief periods in experimental reactors. However, research continues as fusion has some significant potential advantages over fission reactions.

PHYSICSFILE

Cold fusion

Standard fusion processes mimicking those in the Sun are not the only type of fusion reactions being researched. Cold fusion has hit the headlines a number of times over the last 25 years since a couple of scientists falsely claimed to have produced a working cold-fusion process. It is a hypothetical type of nuclear reaction that occurs at near room temperatures, making it potentially more practical than 'hot' fusion which requires temperatures in the millions of degrees. Recent reports suggest that the E-Cat, or Energy Catalyzer, invented by Andrea Rossi may be a step toward making cold fusion a reality—or it may be another hoax.

FIGURE 15.2.6 Italian inventor Andrea Rossi with units of his E-Cat (Energy Catalyzer) cold fusion (Low-Energy Nuclear Reaction, LENR) system.

Fusion as an energy source

Potentially 3.38×10^{14} J of energy could be produced for each kilogram of deuterium–tritium hydrogen fuel compared with 8.80×10^{13} J per kilogram of uranium—almost four times the energy yield per kilogram. Deuterium is an extremely abundant fuel as it can be sourced from the world's oceans.

While some alarmist claims suggest that an out-of-control fusion reactor could turn the Earth into a second Sun, such claims are not based on reality. There is no chain reaction. A fusion reactor requires very little fuel and extremely high operating temperatures. In any breach of the containment vessel, temperatures would quickly cool, extinguishing the fusion reaction.

Fusion reactions also produce very little radioactive waste. Deuterium is not radioactive, nor is the lithium used in 'breeding' tritium. Tritium is radioactive but has an extremely short half-life (12.6 years) when compared with the radioactive waste of fission reactors, and it doesn't need to be transported as uranium and plutonium fuels need to be. It can be produced solely within the reactor itself.

Like fission, fusion produces no greenhouse gases or other atmospheric pollutants. Little radiotoxicity would remain once a fusion power station reached the end of its working life and was decommissioned. What there is would decay to less than levels found in coal-fired power stations in just 100 years. Fuel use is also extremely low compared to conventional and nuclear alternatives. A 1000 MW fusion power station would need just 100 kg of deuterium and three tonnes of lithium per year, compared to 1.5 million tonnes of coal for a similar-sized coal-fired power station. It effectively overcomes many, if not all, of the arguments against fission-powered reactors.

Table 15.2.1 summarises key aspects of the production of power via fission methods versus fusion. The main deterrent to using fusion as a power source is that, at least to date, it remains a highly experimental technology.

	Nuclear fission	Nuclear fusion
Natural occurrence	Fission reactions do not normally occur in nature.	Fusion has been occurring in stars such as the Sun for billions of years.
By-products of the reaction	Fission produces many highly radioactive particles with long half-lives as waste from nuclear reactors.	Few radioactive particles are produced by fusion reactions. Products are only produced in the reactor. Half-life of 12.6 years.
Conditions	A critical mass of the fission fuel and high-energy neutrons are required.	A high-density, high-temperature (at least 40 million K) environment is required.
Energy required to establish process	Takes little energy to split two atoms in a fission reaction.	Extremely high energy is required to bring two small nuclei close enough that nuclear forces overcome their electrostatic repulsion.
Energy released per single event	The energy released by each fission is around 200 MeV.	The energy released by each fusion reaction is around 20–30 MeV.
Energy released per nucleon	Fission involves larger nuclei. The percentage mass defect is lower than for fusion and so the energy per nucleon is lower.	Fusion involves smaller nuclei. The percentage mass defect is greater compared to fission and so the energy per nucleon is greater.
Energy production	Fission is currently used in hundreds of nuclear power plants for generating electricity.	Fusion is an experimental technology for producing power, and large-scale fusion power plants are not expected to operate for many years.
Fuel	Radioactive uranium is the primary fuel used in power plants.	Hydrogen isotopes deuterium (extracted from lakes and oceans) and tritium (produced in the nuclear reactor from lithium) are the primary fuels.

TABLE 15.2.1 A comparison of key aspects of power production through fission and fusion.

FISSION VERSUS THE REST

While fusion may be the hope for almost limitless clean power in the future, fission is currently the only viable source of nuclear power. The major issues in using fission as a power source are well publicised.

Uranium occurs naturally and low-level radioactivity is a natural part of the environment. Supplies of minable uranium could power the nuclear industry for centuries to come. However, the mining of uranium releases radiation in larger amounts directly into the environment surrounding the mine. Processes have to be carefully managed in order to avoid contamination of ground water and water supplies. Australia is a major supplier of uranium to other countries and so mining directly impacts the Australian environment.

A government enquiry into the running of one of Australia's largest uranium mines, the Ranger mine in the Northern Territory (Figure 15.2.7), found the continuation of mining to be high risk and with little long-term promise of financial return. Several radiation leaks highlighted issues in containing radiation to the mine site. Radioactive radon gas posed a serious risk to miners. There have also been concerns that long-term rehabilitation of the site has not been costed or planned. Rehabilitation costs of this one mine have been estimated at several hundreds of millions of dollars.

FIGURE 15.2.7 The open-pit Ranger uranium mine near Kakadu in the Northern Territory. In assessing the environmental impact of nuclear power, each stage of the process from mining to disposal needs to be considered.

Nuclear accidents

Large nuclear accidents of the scale of Chernobyl and Fukushima are rare but very real, and the consequences of these accidents are both widespread and long lasting. Nuclear accidents are rated according to the International Nuclear and Radiological Event scale from to 1 to 7, with 7 being the most serious. Both the Fukushima and Chernobyl disasters were rated at 7.

At the Chernobyl power station (shown in Figure 15.2.8), two workers were killed in the initial explosion and a further 31 died from radiation exposure in the immediate aftermath. However, it is expected that up to a further 6000 people have or may develop thyroid cancers that can be linked to radiation exposure. The entire region around Chernobyl has been evacuated, and the area won't be safe for human habitation for at least another 20 000 years.

FIGURE 15.2.8 The remains of the Chernobyl nuclear power station in the Ukraine soon after the disaster in 1986.

Waste management

The transport and long-term storage of nuclear waste remains an unsolved problem for the nuclear industry. A typical 1000 MW reactor will produce about 25 tonnes of spent fuel rods each year. With more than 400 power stations in operation around the world, this constitutes a very large amount of radioactive waste. By its nature, the waste will need to be safely stored for many thousands of years.

In the US and Canada, spent fuel rods are permanently stored in cooling ponds, as seen in Figure 15.2.9 on page 549, to prevent the release of radioactivity. The U-235 and U-238 nuclei in these fuel rods have half-lives of 700 000 and 4.5 billion years respectively. The numerous fission fragments also have a wide range of very long half-lives. The fuel rods must be permanently stored in cooling ponds inside nuclear power stations in order to protect people from the radioactive emissions. In Japan, Russia and Europe, the spent fuel rods are reprocessed. The uranium is extracted and reused as nuclear fuel, but it's an expensive process.

Australia sent a shipment of intermediate-level waste (ILW) to France for reprocessing more than 10 years ago. This waste was returned in 2015 and temporarily stored at Lucas Heights in Sydney. There is no permanent storage facility in Australia for nuclear waste. There have been plans in the past to build a permanent storage facility, the latest being at Muckaty Station in the Northern Territory. These proposals have always been strongly opposed by people living nearby and no site has ever been constructed. Australia currently has over 4000 m^3 of low- and intermediate-level nuclear waste, and this is growing every year.

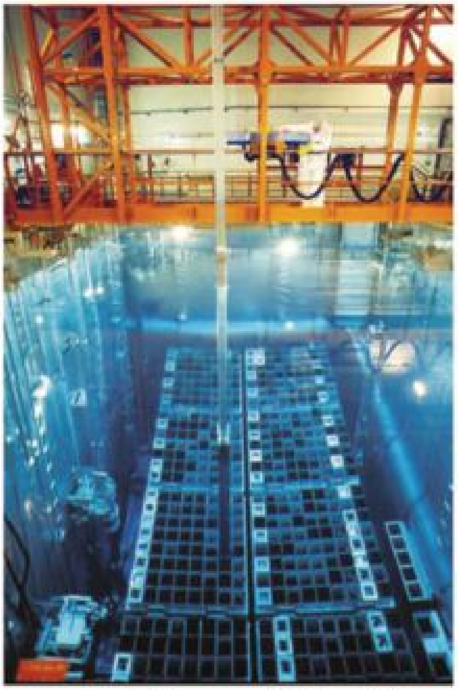


FIGURE 15.2.9 Forty per cent of the world's spent fuel rods are permanently stored in cooling ponds.

Energy decisions

Despite these very significant and real issues, proponents of nuclear power point to nuclear fission as a 'clean' fuel. Greenhouse emissions are negligible and many centuries of fuel remain. In the worst-case scenarios experienced at Chernobyl and Fukushima, the tens of thousands of deaths that were forecast did not occur.

By contrast, it has been estimated that more than 1.2 million people die directly or prematurely as a result of coal mining (see Figure 15.2.10) and airborne pollutants from the coal industry. In China alone, 300 000 people die each year from the effects of the coal industry. Estimates suggest a 10% increase in total health costs in countries with a high dependence on coal. On top of that is the warming of the Earth from increased carbon dioxide levels in the Earth's atmosphere, to which the use of fossil fuel is a significant contributor.

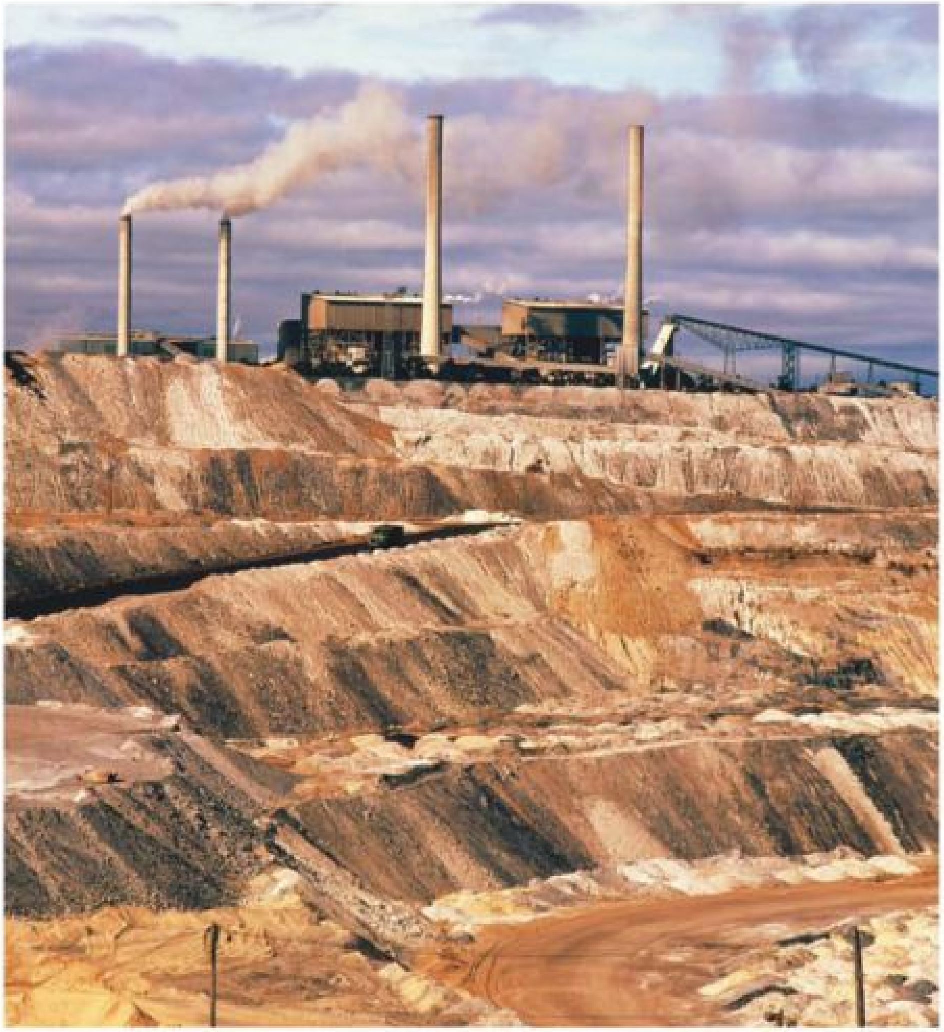


FIGURE 15.2.10 A coal-fired power station and open-cast coal mine in Collie, Western Australia.

These figures are based on many thousands of coal mines and power stations, and ignore the potential for other renewable and sustainable power sources to replace coal without introducing nuclear power.

15.2 Review

SUMMARY

- Only small experimental reactors are testing nuclear fusion as a power source for generators. Every nuclear power station currently in operation and under construction utilises nuclear fission.
- All fission-powered thermal nuclear power stations have several key common features:
  - fuel rods: fissile fuel source
  - moderator: Slow neutrons emitted from fission to increase absorption probability
  - control rods: absorb neutrons to control the rate of chain reaction
  - heat exchanger: exchanges heat energy from the reactor with external systems to create steam
  - generator turbine: produces electricity from the steam.
- In a nuclear reactor, a controlled fission chain reaction occurs to produce thermal energy. The thermal energy from the reactor core is extracted by the coolant and taken to the heat exchanger. This thermal energy is used to produce steam, which drives a turbine and creates electrical energy.
- Fission involves larger nuclei. The percentage mass defect is lower than for fusion and thus the energy per nucleon is lower.
- Radioactive uranium is the primary fuel used in fission power plants. The hydrogen isotopes deuterium (extracted from lakes and oceans) and tritium (produced in the nuclear reactor from lithium) are the primary fuels in fusion.
- Fission power has several large and long-lasting environmental impacts that need to be considered alongside the total environmental impact of conventional coal-fired power stations.

KEY QUESTIONS

1 Name the role or function of the following in a thermal nuclear reactor:

a coolant
b fuel rod
c moderator
d heat exchanger
e control rod.

2 Name the energy transformation that takes place in the following parts of a nuclear power station:

a fuel rods
b heat exchanger
c generator
d power distribution.

3 What are some advantages of fusion reactions compared to fission reactions in terms of energy released?

A Fusion reactions produce more energy per nucleon.
B Fusion reactions create a lot more radioactive waste than fission.
C Each fusion reaction produces more energy than each fission reaction.
D Fusion reactions do not create as much radioactive waste as fission reactions.

4 Without control rods or other moderating mechanisms, what would happen to a fission reaction?

5 Proponents of nuclear power argue that nuclear power is a clean energy. On what is this argument based?

A Greenhouse-gas emissions are negligible.
B Uranium mines have little environmental impact.
C There are no long-term impacts from nuclear power.
D Nuclear power uses much less fuel than alternative sources.

6 A typical 1000 MW fusion-powered station would need just 100 kg of deuterium and three tonnes of lithium per year. How much coal does a similar-sized coal-fired powered station require?

7 One fuel for nuclear fusion is deuterium. What is deuterium and where can it be obtained?

8 Assume that fusion power stations become feasible. What sort of residual radioactivity waste management would be required for the products produced or used once the power station is decommissioned?

9 During the lifetime of a reactor, the control rods need to be gradually raised over a period of months in order to maintain the energy production at a constant rate. Explain why this procedure is necessary.

10 Used nuclear fuel rods from Australia's nuclear reactor at Lucas Heights were shipped to France for reprocessing. These uranium fuel rods were considered to be at no risk of exploding like a nuclear weapon. What does this suggest about the proportion of fissile uranium in the fuel rods?

Chapter review

15

KEY TERMS

actinides
chain reaction
control rods
coolant
core
critical mass
fast reactor
fuel rods
heat exchanger
heavy water
moderator
nuclear fission reactor
nuclear power station
subcritical mass
supercritical mass

1 Which one of the following explains why it is easier to trigger nuclear fission with a neutron rather than an alpha particle?

A Neutrons are more massive than alpha particles.
B Neutrons have more energy than alpha particles.
C Neutrons are uncharged and are not repelled by the nucleus.
D Neutrons are not impeded by electrons but alpha particles are.

2 Uranium-238 is effectively non-fissile.

a What does non-fissile mean?
b How is uranium-238 used in fast reactors?
c What advantage is there in using uranium-238 in a nuclear reactor?

3 The fuel rods that are used in a thermal nuclear reactor contain:

A mostly uranium-235
B mostly uranium-238
C about 50% uranium-235 and 50% uranium-238
D mostly uranium ore (yellowcake)

4 Complete the following sentences about fission by choosing the correct option from those provided in the brackets.

The total numbers of protons and [**electrons/neutrons/positrons**] in the fragments from the fission of a U-235 nucleus are [**lost/conserved/split**]. The small [**charge/fission/mass defect**] will be equivalent to the [**energy/temperature/power**] released.

5 Determine the number of neutrons *X* released as a result of this fission reaction:

$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{140}_{54}Xe + ^{94}_{38}Sr + x^{1}_{0}n$

6 A typical fission of uranium is:

$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{141}_{56}Ba + ^{92}_{36}Kr + 3^{1}_{0}n$

Explain why the incident neutron should be 'slow' for fission to occur.

7 Uranium ore that is mined from the ground contains uranium-235 and uranium-238.

a Which of these isotopes is the more prevalent?
b Which of these isotopes is the more fissile?
c Explain why a chain reaction does not occur in uranium ore.

8 Discuss how the chain reaction that releases the energy in a nuclear fission explosion is established. Explain why the chain reaction grows rather than dies out.

9 The critical mass of uranium-235 is about 1 kg. Explain why a 5 kg piece of U-235 that is flattened into a sheet is not capable of exploding.

10 Explain the effect on a nuclear fission chain reaction if the average number of neutrons per fission continuing the chain reaction is:

a less than one
b equal to one
c greater than one.

11 Briefly describe the techniques that are currently being used in tokamak-style nuclear-fusion reactors.

12 The principle process responsible for the energy output from the Sun is believed to be from a proton–proton cycle in which four protons are combined to form one helium nucleus. Write down the proton–proton cycle reaction.

13 **a** Outline the process by which a nuclear power plant produces electricity.
b Discuss the primary difference between the way electricity is produced in a nuclear power station and in a coal-burning power station.
c What aspects of the production of electricity do coal-fired and nuclear power stations have in common?

14 Name the function of the following parts of a nuclear reactor:

a control rod
b radiation protection barrier
c moderator
d cold coolant
e fuel rod
f hot coolant.

15 One kilogram of uranium-235 is capable of releasing 6.8×10^{13} J of energy during nuclear fission. In comparison, the burning of 1 kg of coal releases around 2.5×10^{7} J. Calculate the number of tonnes of coal that is burnt to provide the same energy as that released by 1 kg of uranium-235.

Chapter review *continued*

16 When nuclear fission takes place in the core of a nuclear reactor, a great quantity of heat energy is produced. How is this heat energy converted into electrical energy? Choose the correct answer from those provided in the brackets.

Heat from the [**enrichment/fission/generator**] process is carried by the [**coolant/control rods/fuel rods**] to the heat exchanger. This heat energy is used to make [**fission/steam/uranium**] and drive the turbine and generator to produce electricity.

17 Briefly discuss the role of the moderator in a thermal nuclear reactor. List two materials that are used as moderators.

18 Briefly discuss and compare fission and fusion in terms of how much energy is released and where this energy comes from.

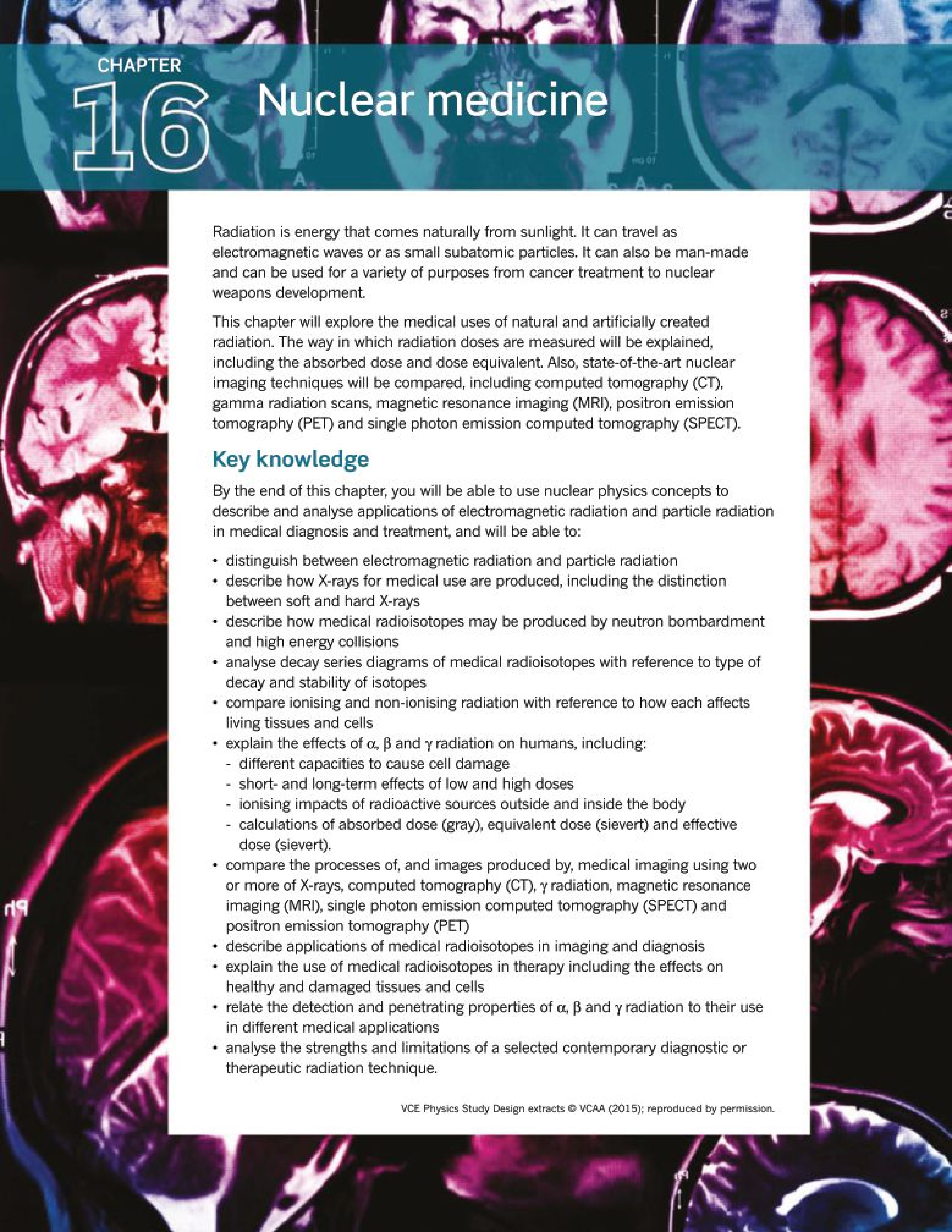
19 Fusion reactors potentially have significant advantages over fission reactors, but there are also disadvantages. Summarise the main advantages and disadvantages covered in this chapter.

20 Australia sent a shipment of intermediate-level waste to France for reprocessing more than 10 years ago. This was returned to Australia in 2015 to be temporarily stored at Lucas Heights in Sydney. Research the following:

a the classification criteria for waste to be rated as intermediate-level waste (ILW)

b what happened to the waste during this reprocessing process

c what the current levels of waste are in Australia.



CHAPTER 16 Nuclear medicine

Radiation is energy that comes naturally from sunlight. It can travel as electromagnetic waves or as small subatomic particles. It can also be man-made and can be used for a variety of purposes from cancer treatment to nuclear weapons development.

This chapter will explore the medical uses of natural and artificially created radiation. The way in which radiation doses are measured will be explained, including the absorbed dose and dose equivalent. Also, state-of-the-art nuclear imaging techniques will be compared, including computed tomography (CT), gamma radiation scans, magnetic resonance imaging (MRI), positron emission tomography (PET) and single photon emission computed tomography (SPECT).

Key knowledge

By the end of this chapter, you will be able to use nuclear physics concepts to describe and analyse applications of electromagnetic radiation and particle radiation in medical diagnosis and treatment, and will be able to:

- distinguish between electromagnetic radiation and particle radiation
- describe how X-rays for medical use are produced, including the distinction between soft and hard X-rays
- describe how medical radioisotopes may be produced by neutron bombardment and high energy collisions
- analyse decay series diagrams of medical radioisotopes with reference to type of decay and stability of isotopes
- compare ionising and non-ionising radiation with reference to how each affects living tissues and cells
- explain the effects of α, β and γ radiation on humans, including:
  - different capacities to cause cell damage
  - short- and long-term effects of low and high doses
  - ionising impacts of radioactive sources outside and inside the body
  - calculations of absorbed dose (gray), equivalent dose (sievert) and effective dose (sievert).
- compare the processes of, and images produced by, medical imaging using two or more of X-rays, computed tomography (CT), γ radiation, magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT) and positron emission tomography (PET)
- describe applications of medical radioisotopes in imaging and diagnosis
- explain the use of medical radioisotopes in therapy including the effects on healthy and damaged tissues and cells
- relate the detection and penetrating properties of α, β and γ radiation to their use in different medical applications
- analyse the strengths and limitations of a selected contemporary diagnostic or therapeutic radiation technique.

16.1 Producing medical radiation

FIGURE 16.1.1 Roentgen's first radiograph of his wife's hand taken in 1895 shortly after his discovery of X-rays. The bulge on one of her fingers is a ring that she is wearing.

X-rays are still the most common form of medical imaging. Since their discovery in 1895 by Wilhelm Roentgen (see Figure 16.1.1), they have been widely used for over one hundred years for this purpose. For example, in the Balkan War of 1897, X-rays were used to locate bullets in wounded soldiers.

However, the effects of overexposure to X-rays were not known at that time, leading to many people receiving excessive doses of radiation. This unawareness of the hazards of radiation led many scientists who worked with X-rays to suffer hair loss, burns, skin ulcers, skin cancers and death from the unprotected overuse of X-rays. The way in which X-rays and other medically relevant forms of radiation are created, as well their radioactive decay over time, is the focus of this section.

TYPES OF RADIATION

The term 'radiation' is widely used and at times very poorly understood. There are many different forms of radiation, and the degree of danger that they present depends on their ability to interact with atoms. All life is exposed to a variety of radiation during every second of its existence. Radiation is made up of a range of energies with different wavelengths and frequencies. Together they form the electromagnetic spectrum. The electromagnetic spectrum was described in detail in Chapter 2.

Within this spectrum, low-energy electromagnetic radiation is commonly referred to as **non-ionising radiation**. This is because it does not carry enough energy to break molecular bonds and displace (remove) electrons from atoms. This radiation has very low frequency (below 1–2 × 10^{15} hertz), very low energy and long wavelengths (longer than 100 nm). It includes radio waves, microwaves, visible light and UV-A radiation (see the left-hand side of Figure 16.1.2). All living things are exposed to significant amounts of non-ionising radiation every day without serious consequences.

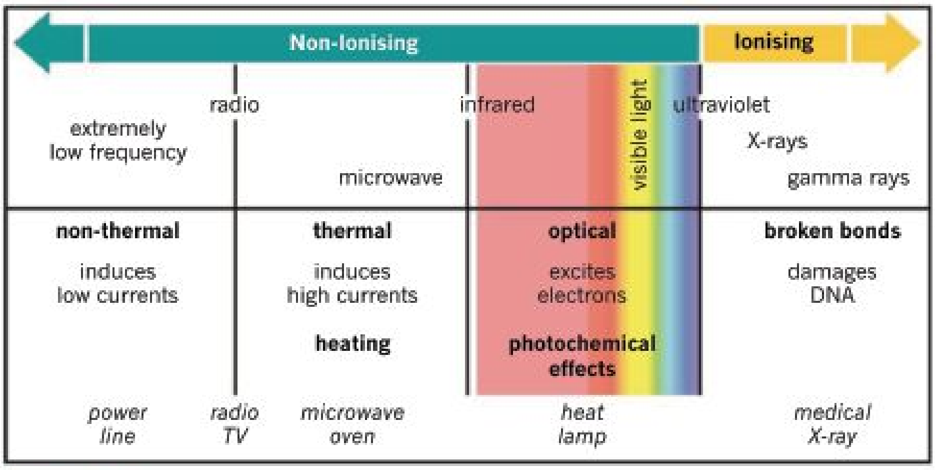


FIGURE 16.1.2 Ionising and non-ionising radiation in the electromagnetic spectrum and their respective effects when interacting with atoms.

There is also radiation that has very high energy levels, enough to break apart molecular bonds within molecules and to remove outer-shell electrons from atoms. This alters the atom's original structure and creates ions. For this reason, this type of radiation is known as ionising radiation. Electromagnetic radiation with a frequency above 2 × 10^{16} Hz is ionising. Thus gamma rays, X-rays and ultraviolet B and C radiation are ionising radiation (see the right-hand side of Figure 16.1.2).

When ionising radiation interacts with living tissues, it may induce genetic defects. This is due to its ionising effect on certain biomolecules (mostly nucleic acids and proteins) that are critical for cell survival. Alteration to these biomolecules will lead to significant biological damage. By damaging the bonds in the **DNA** and/

or the structure of proteins that are critical for cellular growth, ionising radiation may destroy the affected cells. It may also induce the development of cancerous tumours from the radiated cell population.

Alpha particles, beta particles, gamma rays and X-rays are all examples of ionising radiation. Recall from Chapter 6 that some radiation, such as α and β radiation, come in the form of tiny particles of matter. These particles are expelled from the nuclei of unstable atoms. For this reason, ionising radiation is also called **particle radiation**. However, γ and X-rays are high-energy electromagnetic waves, consisting of high-energy photons (light) with no mass.

In summary, radiation can be classified as being either electromagnetic or particle, as shown in Figure 16.1.3.

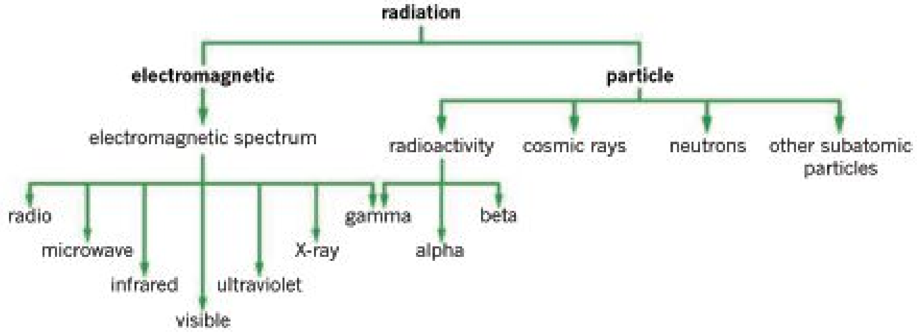


FIGURE 16.1.3 Diagram showing electromagnetic and particle radiation.

Types of X-rays

Wilhelm Roentgen's discovery of X-rays at the end of the nineteenth century transformed medicine practically overnight. For the first time ever in history, the inner details and structure of the human body could be seen without cutting into it.

Soft and hard X-rays

There are two types of X-rays, which differ in wavelength and penetrating ability:

- **soft X-rays**: X-rays with wavelengths greater than 0.1 nm
- **hard X-rays**: X-rays with wavelengths less than 0.1 nm.

Hard X-rays have greater energy (shorter wavelength) and are more penetrating than soft X-rays. They are also less absorbed by matter than soft X-rays. The characteristics of hard and soft X-rays are shown in Table 16.1.1.

Hard X-rays	Soft X-rays
wavelength <0.1 nm	wavelength >0.1 nm
higher frequency	lower frequency
greater penetrative capacity	inferior penetrative capacity
high voltages used in their production	low voltages used in their production
superior resolution of radiographs	poorer resolution of radiographs

TABLE 16.1.1 Comparison of hard and soft X-rays.

X-RAY PRODUCTION

X-rays that are used for medical imaging procedures and research are produced by accelerating a beam of electrons with a high voltage along a vacuum tube and allowing them to collide with a metal target. The electrons then slam into a positively charged target (usually made of tungsten), losing their energy very rapidly. X-rays are produced as a result.

In this way, X-rays result from the conversion of the kinetic energy attained by the electrons accelerated under a high voltage into electromagnetic radiation, as a result of high-speed collisions. Essential instruments in this process are an X-ray tube and X-ray generator.

PHYSICSFILE

Wilhelm Roentgen

Wilhelm Roentgen (1845–1923) was a German physicist. He studied engineering in Holland and in 1888 he became a professor at the Institute of Physics in Wurzburg in Germany. In 1895, Roentgen noticed that a cathode lamp made a barium-coated piece of paper shine from a distance. The lamp emitted a type of radiation that was not known at that time, and which caused the glowing effect. Roentgen named the radiation X-rays, as X stands for an unknown. His discovery, for which he was awarded the Nobel Prize in 1901, revolutionised medicine, and made it possible to take a look inside a human body. Roentgen refused to have the invention, or its use, patented.

FIGURE 16.1.4 Wilhelm Roentgen (1845–1923).

The X-ray tube provides the suitable environment and components needed to generate X-rays. X-ray tubes always include three main components: the heated filament, an accelerating potential (high voltage) and a target metal. Voltages of around 25 kV are used to produce soft X-rays. A diagram of an X-ray tube is shown in Figure 16.1.5. The X-ray generator provides the electrical power that is needed for the reaction to occur.

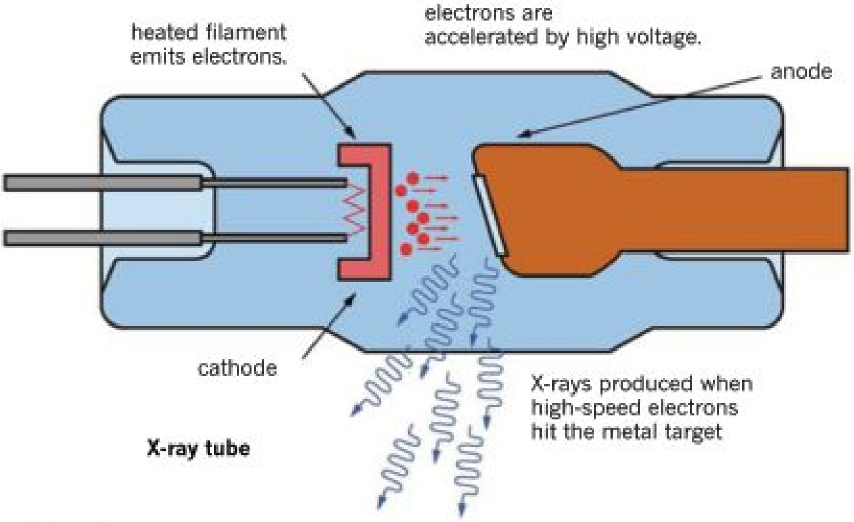


FIGURE 16.1.5 A diagram of an X-ray tube.

Production of soft X-rays

Soft X-rays that are used for medical imaging are produced by the same principle of accelerating a beam of electrons along a vacuum tube as described above. Voltages of typically 25 000 V give the electrons a high enough speed to form X-rays with a frequency of 1–2 × 10^{16} Hz (see Figure 16.1.6).

FIGURE 16.1.6 X-ray machine used to create soft X-rays. The soft X-rays will generate a jaw X-ray image of this patient.

Production of hard X-rays

Hard X-rays have more energy and so greater penetrating ability than soft X-rays. They are produced by using extremely high accelerating voltages of between 4 and 25 million volts (4–25 MV). The principle of X-ray production from rapidly decelerated electrons still applies to hard X-rays. However, due to the extremely high voltages used, these X-ray machines have a slightly different set-up from those used for the production of soft X-rays.

Linac (an abbreviation of 'linear accelerator') machines are used to produce hard X-rays. Parts of these machines can rotate so that the beam of X-rays can be sent into the patient from different directions as shown in Figure 16.1.7.

FIGURE 16.1.7 Hard X-rays created by this linear accelerator machine (linac) are being used to destroy a cancer tumour in a patient.

The hard X-rays used for radiation therapy have a much higher frequency and are much more energetic than those used to produce diagnostic images. They are highly penetrating and ionising and so the room in which the therapy takes place needs to have thick concrete walls and lead lining to protect others from unwanted exposure to the radiation.

PHYSICS IN ACTION

Shoe-fitting X-ray machines

If your grandparents or great grandparents are still alive, they may remember when shoe shops sometimes had a small X-ray machine (Figure 16.1.8). Customers would place their feet and shoes inside the machine and an X-ray would be taken. They were commonly used with children so parents could see if there was room for the child's feet to grow in the shoes. These machines were banned about sixty years ago due to rising public health concerns.

FIGURE 16.1.8 In the first half of the twentieth century, some shoe shops had X-ray machines that were used to see if a shoe was a good fit. These machines would have produced soft X-rays.

PRODUCTION OF RADIOISOTOPES

Naturally occurring radioisotopes

Many of the chemical elements found in nature have a number of isotopes which have different masses due to different numbers of neutrons in their nuclei. However, these isotopes retain the same number of protons. When an isotope contains an unstable combination of neutrons and protons, the resulting atom is called a radioisotope. The unstable nucleus decays emitting α, β or γ radiation until stability is reached over a particular length of time. This unstable combination of neutrons and protons can occur spontaneously in nature, (for example in radium-226) or by artificially altering the atoms of an element (for example in cobalt-60).

Man-made radioisotopes: Artificial transmutation

Mostly naturally occurring radioisotopes were used in the early days of medical research. Today, most of the radioisotopes that are used in medical applications are created by a process known as **artificial transmutation** that uses **neutron bombardment**. Neutron bombardment can take place inside a nuclear reactor or in a particle accelerator device called a **cyclotron**. Some hospitals have their own cyclotrons to make radioisotopes with short half-lives.

The principle behind neutron bombardment is quite simple. A sample of a stable isotope is placed inside a nuclear reactor and bombarded with neutrons, alpha particles, or other subatomic particles. When high-energy collisions occur between the bombarding, or irradiating, particles and the nucleus of the stable isotope, it creates an unstable isotope of the sample element (see Figure 16.1.9).

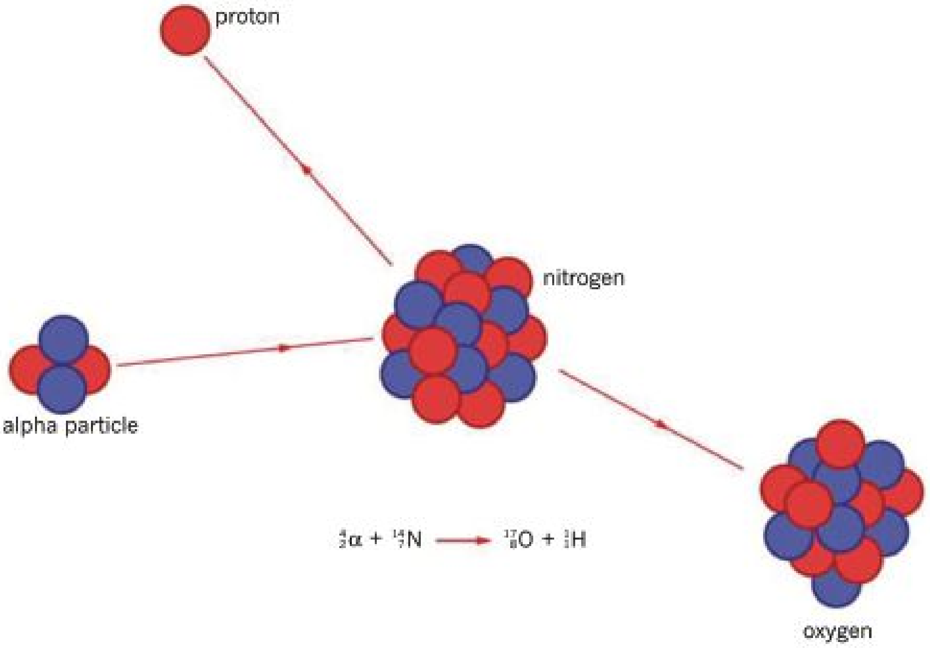


FIGURE 16.1.9 Artificial transmutation of nitrogen atoms into oxygen and hydrogen (a proton). In his famous experiment in 1919, Ernest Rutherford bombarded atoms of gaseous nitrogen with alpha particles emitted from a piece of radioactive ore, which led him to find small amounts of oxygen and hydrogen.

For example, the medical radioisotope cobalt-60 (widely used for cancer treatment) is manufactured via artificial transmutation. A sample of the naturally occurring and stable isotope cobalt-59 is irradiated with neutrons. Some of the cobalt-59 nuclei absorb neutrons, and this results in a quantity of cobalt-60 being produced, as shown in Figure 16.1.10.

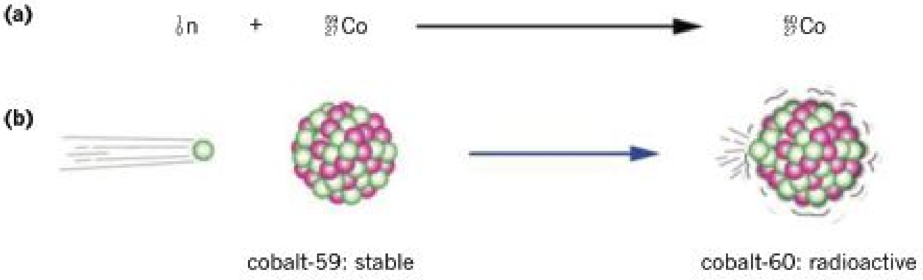


FIGURE 16.1.10 The artificial radioisotope cobalt-60 is produced by bombarding a sample of naturally occurring cobalt-59 with neutrons in a process known as artificial transmutation. (a) The chemical equation describing the formation of cobalt-60 (b) An Illustrative diagram of the particles involved in the formation of cobalt-60.

Other examples of medical radioisotopes that are commonly produced by neutron bombardment include:

(a) Molybdenum-99: a sample containing uranium is irradiated according to the following reaction:

$$^{235}_{92}U + ^{1}_{0}n \xrightarrow{\text{fission}} ^{99}_{42}Mo + ^{134}_{50}Sn + 3^{1}_{0}n + \gamma + \text{heat}$$

The decay of Mo-99 generates one of the most widely used medical diagnostic radioisotope worldwide, Tc-99m.

$$\begin{array}{lll} ^{99}_{42}Mo \xrightarrow{\beta} & ^{99m}_{43}Tc \rightarrow & ^{99}_{43}Tc + \gamma \\ t_{½} = 66\text{ h} & t_{½} = 6\text{ h} & t_{½} = 2.1 \times 10^{5}\text{ years} \end{array}$$

Tc-99m is a highly versatile medical tracer. It is used to examine the functioning of several organs of the body, including the brain, heart, kidneys, bones, bone marrow and stomach. It can also show the presence of cancerous tumours.

(b) Iodine-131: also called 'radioactive iodine'. This radioisotope has a short half-life of about 8.0 days, and it is highly radioactive. Its mode of β and γ radiation decay is known for its high toxicity, causing mutations and death of the cells it penetrates and their surrounding area. Inside the body it has the tendency to accumulate in the thyroid gland, making it very useful in the treatment and diagnosis of thyroid conditions.

$$\begin{array}{l} ^{130}_{52}Te + ^{1}_{0}n \rightarrow ^{131}_{52}Te(+\gamma) \rightarrow ^{131}_{53}I + ^{0}_{-1}\beta + \gamma \\ \qquad\qquad\qquad t_{½} = 25\text{ min} \quad t_{½} = 8.02\text{ days} \end{array}$$

The decay of $^{131}_{53}I$ follows the following formula:

$$^{131}_{53}I \rightarrow ^{131}_{54}Xe + ^{0}_{-1}\beta + \gamma$$

The negative β particles have high energy and thus are very effective at destroying cancer cells. After administration of I-131 to a patient, the person is asked to remain in isolation due to possible radiation contamination of their environment. The patient should not come in contact with pregnant women or young children because the strong γ emissions could cause genetic damage to unborn babies or infants.

PHYSICS IN ACTION

Discovery of artificial radioactivity

Following in the footsteps of her parents, Marie and Pierre Curie, French scientist Irene Joliot-Curie discovered artificial radioactivity, together with her husband Frederic Joliot-Curie (pictured in Figure 16.1.11).

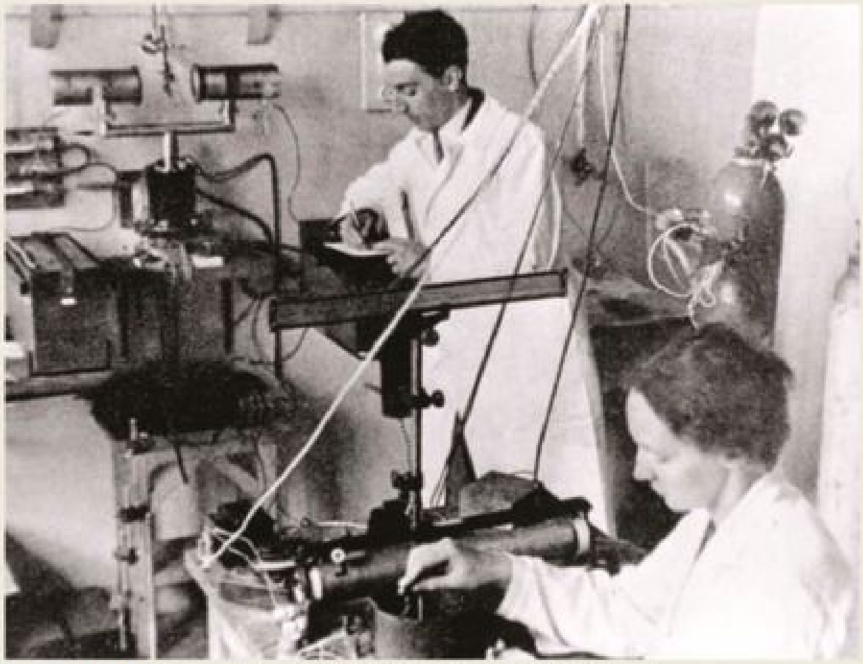


FIGURE 16.1.11 Irene and Frederic Joliot-Curie at work in their laboratory.

Irene's contribution and commitment to the advancement of nuclear physics were monumental. The Joliot–Curie team generated the first artificial radioactive element made from stable atoms. They used α particles (helium nuclei, He^{2+}) to bombard various stable elements, including aluminium and boron. As a result, they obtained radioactive isotopes from those elements. This discovery not only added new radioisotopes to the periodic table of the time but also made those radioisotopes available for further scientific and medical applications.

For this work the Joliot-Curies were awarded the Nobel Prize in Chemistry in 1935, 'in recognition of their synthesis of new radioactive elements'.

Besides her scientific work, Irene had a keen interest in the social and intellectual development of women. She served as a member of the Comité National de l'Union des Femmes Francais, and the World Peace Council.

Production of medical radioisotopes in Australia

Australia is one of the few counties that produce clinically relevant radioisotopes. Currently, nearly 2 million potential doses of isotopes are produced each year. These are used to diagnose and treat different types of cancers, skeletal injuries and heart disease.

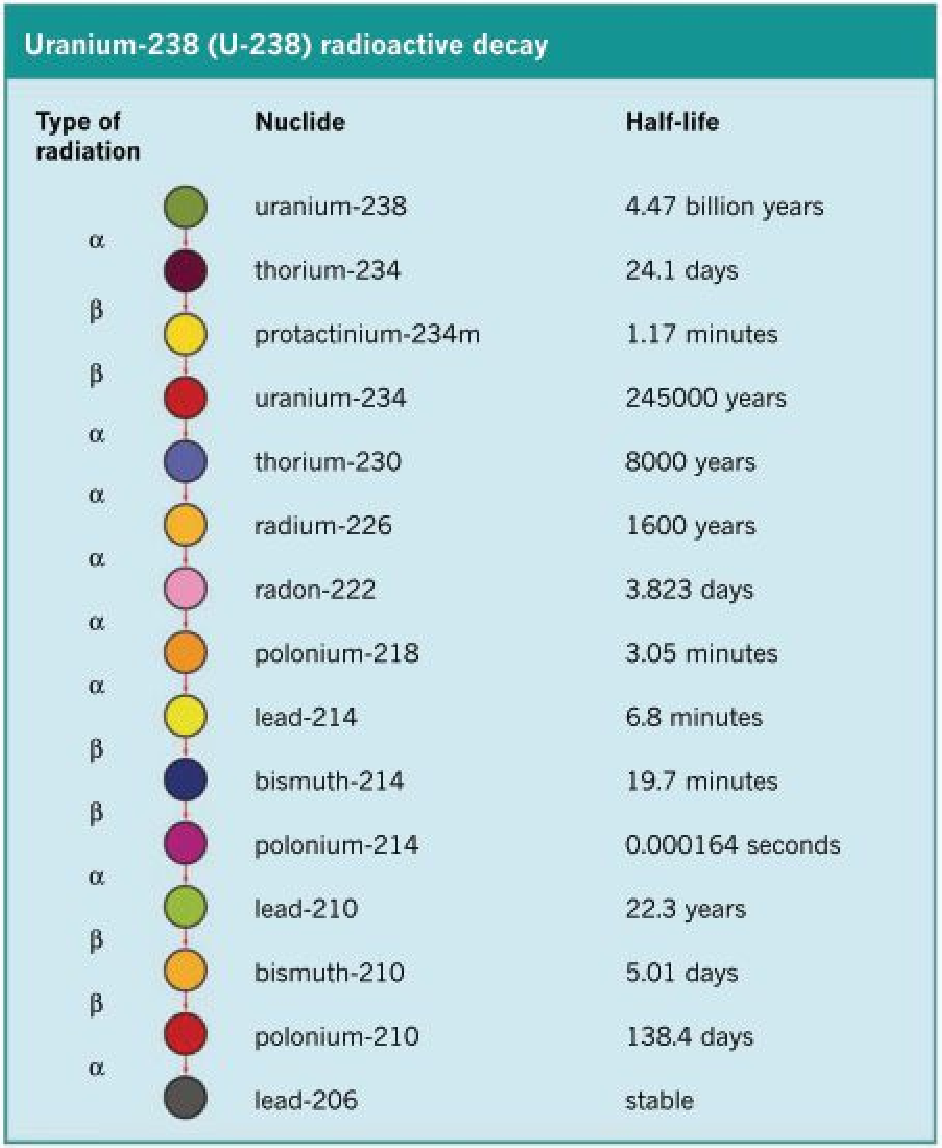
The production of these medicines takes place at the Australian Nuclear Science and Technology Organisation (ANSTO) facility in New South Wales, which is pictured in Figure 16.1.12. The most common produced radioisotope is Mo-99, which is distributed around the country and, when needed, around the world.

DECAY SERIES OF RADIOISOTOPES

Often when a radioactive nucleus decays, the newly created nucleus is itself also radioactive; therefore, a new decay event needs to occur until a non-radioactive and stable atomic nucleus is formed.

This leads to the formation of what is called a **decay series**, which charts the decay process until a stable nucleus is reached (see Figure 16.1.13). Each different radioisotope in the series has its own half-life or $t_{\frac{1}{2}}$. In general, the shorter the half-life of an atom, the faster it will disintegrate and the more active the element.

FIGURE 16.1.12 The ANSTO Nuclear Medicine facility in New South Wales.



Uranium-238 (U-238) radioactive decay

Type of radiation	Nuclide	Half-life
	uranium-238	4.47 billion years
α		
	thorium-234	24.1 days
β		
	protactinium-234m	1.17 minutes
β		
	uranium-234	245000 years
α		
	thorium-230	8000 years
α		
	radium-226	1600 years
α		
	radon-222	3.823 days
α		
	polonium-218	3.05 minutes
α		
	lead-214	6.8 minutes
β		
	bismuth-214	19.7 minutes
β		
	polonium-214	0.000164 seconds
α		
	lead-210	22.3 years
β		
	bismuth-210	5.01 days
β		
	polonium-210	138.4 days
α		
	lead-206	stable

FIGURE 16.1.13 The natural decay series of uranium-238.

Uranium-238 has a very long decay series and undergoes fourteen decays until it forms lead-206 (Pb-206), which is a stable element. Uranium ore is a mix of many different radioisotopes, each one with its own half-life. Over time, the amount of Pb-206 increases. Medical radioisotopes usually have much shorter decay series, and thus they can be represented with a few sets of equations.

16.1 Review

SUMMARY

- Radiation can exist as electromagnetic radiation or particle radiation. Electromagnetic radiation comes in the form of photons (light) with no mass, while particle radiation consists of subatomic particles with high energy.
- Non-ionising radiation includes radio waves, microwaves, visible light and UV-A radiation. All living things are exposed to significant amounts of such radiation every day without serious consequences.
- X-rays can be classified as soft or hard X-rays. Soft X-rays have wavelengths larger than 0.1 nm, while hard X-rays have much shorter wavelengths of less than 0.1 nm.
- Hard X-rays have greater energy and are more penetrating than soft X-rays.
- X-rays are produced by the collision of accelerated electrons, emanating from a heated filament, into a metal target.
- Radioisotopes can be naturally derived or synthetised artificially. Natural radioisotopes are mostly used for industrial applications, while artificial ones have wide application in medical diagnosis.
- The process of changing one element into another is called transmutation. Artificial transmutation is achieved by neutron or alpha particle (or other subatomic particle) bombardment of an atomic nucleus at high speed inside a nuclear reactor. This creates an unstable isotope of the sample element.

KEY QUESTIONS

1 List the seven forms of electromagnetic radiation in order from lowest to highest energy.

2 Your dentist takes an X-ray of your tooth. Is the radiation being used ionising or non-ionising and is it hard or soft X-rays?

3 A radiologist uses a linac machine to deliver a dose of X-rays to a cancer patient to destroy a tumour. Is the radiation being used ionising or non-ionising and is it hard or soft X-rays?

4 The artificial transmutation of cobalt-59 takes place in the core of a nuclear reactor. To create the new radioisotope cobalt-60, what is the parent nucleus bombarded with?

5 Molybdenum $^{99}_{42}$Mo is a beta emitter with a half-life of 67 hours. It decays to form metastable technetium, $^{99m}_{43}$Tc, which has a half-life of 6 hours. This in turn decays to form technetium $^{99}_{43}$Tc, which has a half-life of around 250 000 years.

a Which is the least stable of the three radioisotopes?

b Which is the most stable of the three radioisotopes?

6 Compare the ionising ability and wavelengths of hard and soft X-rays.

7 Radioactive iodine for medical purposes is produced by neutron bombardment of tellurium atoms. Why is iodine-131 so useful for cancer treatment?

8 Give an example of a natural radioisotope with a long decay series.

9 Compare the medical uses of hard and soft X-rays.

10 What energy conversion takes place when X-rays are generated?

16.2 Measurement of radiation doses

Ionising radiation from alpha, beta and gamma sources is harmful to humans and other living things. People who work with radiation in fields such as medicine, mining, nuclear power plants and industry must be able to closely monitor their radiation dose (see Figure 16.2.1). Similarly, radiologists who administer courses of radiation treatment to cancer patients need to be able to measure the radiation dose that they are applying. In this section, radiation doses and how they are measured will be explained.

FIGURE 16.2.1 A dosimeter used to monitor gamma radiation exposure.

THE EFFECTS OF RADIATION ON LIVING ORGANISMS

If possible, exposure to ionising radiation should be avoided. When alpha, beta or gamma radiation or an X-ray pass through a body cell, it may turn molecules in the cell into an ion pair. For example, if the radiation ionises a water molecule then a hydrogen ion and a hydroxide ion can be formed. This is shown in Figure 16.2.2. These ions are highly reactive and can attack the DNA that forms the chromosomes in the nucleus of the cell. This may cause the cell either to die or to divide and reproduce at an abnormally rapid rate. When the latter occurs, a cancerous tumour may form.

FIGURE 16.2.2 Ionising radiation has enough energy to break the bonds within a water molecule and create a pair of ions.

The effects of a dose of ionising radiation can be divided into two groups: the short-term **somatic** effects and the long-term **genetic** effects.

Somatic effects

Somatic effects arise when ordinary body cells are damaged, and depend on the size of the dose. Very high doses lead to almost immediate symptoms; lower doses could lead to symptoms developing many years later. Table 16.2.1 outlines the somatic effects of radiation doses.

Whole body dose (Sv)	Symptom
<1	non-fatal only minor symptoms such as nausea white blood cell level drops
2	death unlikely radiation sickness, i.e. nausea, vomiting and diarrhoea skin rashes hair loss bone-marrow damage
4	50% likelihood of death within 2 months severe radiation sickness high probability of leukaemia and tumours
8	almost certain death within 1 or 2 weeks acute radiation sickness—convulsions, lethargy

TABLE 16.2.1 The somatic effects of radiation doses.

Genetic effects

When cells in the reproductive organs (ovaries or testes) are damaged, the body suffers genetic effects of ionising radiation. Cells in the reproductive organs develop into ova and sperm, so if the DNA in the chromosomes of these cells is damaged, this genetic change could be passed on to a developing embryo. The DNA changes in these damaged cells are known as **mutations**.

There are many different ways in which genetic defects can show up in future generations. Possible defects include poor limb development, harelips and other birth abnormalities. They may surface in the next generation or lie dormant for several generations. In other words, if a person suffers damage to their reproductive cells, their children may seem to be unaffected but their grandchildren may be genetically weakened.

FIGURE 16.2.3 A cancer patient about to receive radiation therapy. Medical staff would take precautions to protect the patient's reproductive organs before exposure to ionising radiation.

For these reasons, when a patient is undergoing radiotherapy, it is most important that their reproductive organs are well shielded from the radiation (see Figure 16.2.3). These organs are among the most radiosensitive organs (i.e. easily damaged by radiation) in the body. Patients who have not yet started a family may have eggs removed from ovaries or sperm samples taken and frozen for later use.

A developing foetus is also very sensitive to radiation and so pregnant women should avoid having X-rays. For this reason foetal images are gathered using ultrasound techniques.

MEASURING RADIATION EXPOSURE

Absorbed dose

Exposure to high-energy radiation such as a radioactive source or an X-ray is harmful to living tissue. The energy of the radiation acts to break apart molecules and ionise atoms in the body's cells. The severity of this exposure depends on the amount of radiation energy that has been absorbed and the mass of tissue involved.

The radiation energy absorbed per kilogram of tissue is called the **absorbed dose**.

$$\text{Absorbed dose} = \frac{\text{energy absorbed by tissue}}{\text{mass of tissue}}$$

The absorbed dose is measured in joules per kilogram (J kg^{-1}) or grays (Gy), i.e. 1 Gy = 1 J kg^{-1}.

Dose equivalent

The absorbed dose is not widely used when measuring the radiation dose. That's because it does not take into account the type of radiation involved. **Dose equivalent**, which does take the type of radiation involved into account, is the most common way in which radiation doses are measured.

Alpha particles are the most ionising form of radiation. This is because of their relatively low speed, high charge and large mass. Alpha particles interact with, and ionise, almost every atom that lies in their path. This means that an absorbed dose of alpha radiation is about 20 times more damaging to human tissue than an equal absorbed dose of beta or gamma radiation.

The weighting of the biological impact of radiation is called the **quality factor (QF)** or weighting factor. A list of quality factors is shown in Table 16.2.2.

Radiation	Quality factor
α-particles	20
neutrons* (10 keV)	10
β-particles	1
γ-rays	1
X-rays	1

*Radiation from neutrons is only found around nuclear reactors and neutron bomb explosions.

TABLE 16.2.2 Quality factors for different types of radiation.

Gamma rays and X-rays have relatively low ionising power. They have no charge and move at the speed of light, so they fly straight past most atoms and interact only occasionally as they pass through a material or substance. Gamma rays and X-rays would cause only slight damage to living cells. Beta-minus particles (electrons) are considered to be as damaging as gamma rays and X-rays. This is reflected in their low quality factor of 1, as shown in Table 16.2.2.

Neutrons are a damaging form of radiation and have a quality factor of 10. This is why accidents at nuclear reactors are so dangerous to workers who are present.

The dose equivalent takes into account the absorbed dose and the type of radiation. This gives a more accurate picture of the actual effect of the radiation on a person.

Dose equivalent = absorbed dose × quality factor

$$DE = AD \times QF$$

The dose equivalent is measured in sieverts (Sv).

An absorbed dose of just 0.05 Gy of alpha radiation is equally as damaging to a person as an absorbed dose of 1.0 Gy of beta radiation. While less energy is carried by the α-particles than the β-particles, each α-particle does far more damage. In each case, the dose equivalent is 1 Sv, and 1 Sv of any radiation causes the same amount of damage (see Table 16.2.1 on page 563).

Background radiation

It is important to appreciate that 1 Sv is a massive dose of radiation. It would not be fatal, but it would certainly lead to a severe case of radiation sickness. In Australia, the average annual background radiation dose is about 2.0 mSv, or 2000 μSv. Using Table 16.2.3, which outlines annual radiation doses in Australia, it is possible to estimate your dose over the past year.

Radiation source	Average annual dose (μSv)	Local variations
cosmic radiation	300	plus 200 μSv for each round-the-world flight plus 20 μSv for each 10° of latitude (Melbourne is at 37.8° latitude) plus 150 μSv if you live 1000 m above sea level
rocks, air and water	1350	plus 1350 μSv if you live underground plus 1350 μSv if your house is made of granite minus 140 μSv if you live in a weatherboard house
radioactive food and drink	350	plus 1000 μSv if you have eaten food affected by the Fukushima fallout
manufactured radiation	60	plus 60 μSv if you live near a coal-fired power station plus 30 μSv from nuclear testing in the Pacific
medical exposure	variable	plus 30 μSv for a chest X-ray plus 300 μSv for a pelvic X-ray plus 5000 μSv for a CT scan plus 40 000 000 μSv for a course of radiotherapy using cobalt-60

TABLE 16.2.3 Radiation doses from different sources

PHYSICSFILE

Share of the different sources of radiation

The average equivalent of the dose absorbed by a person in a year is equal to a few millisieverts.

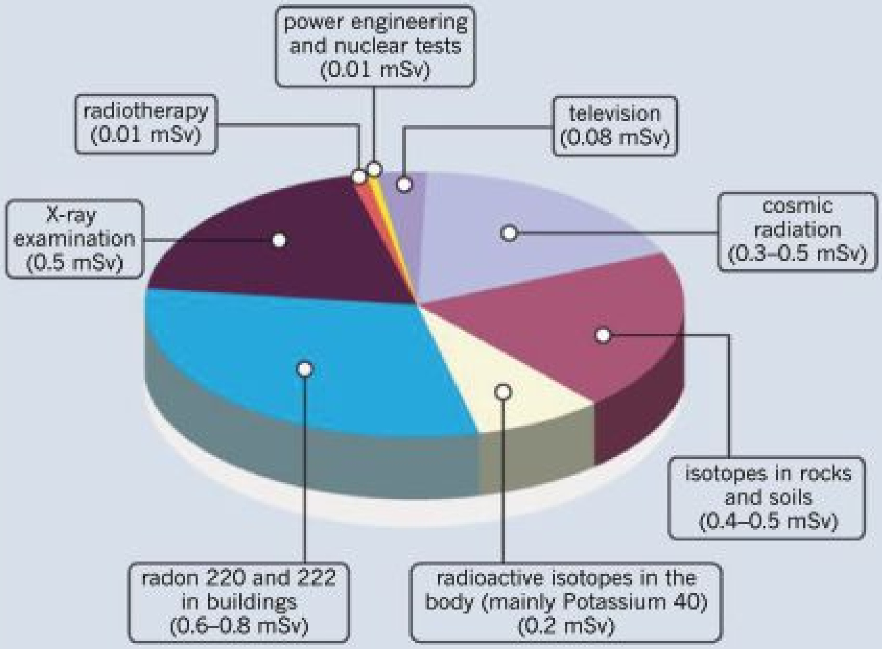


FIGURE 16.2.4 The average equivalent of the dose absorbed in a year from natural and artificial sources.

PHYSICSFILE

The value of dose equivalent

In India there is a village which is situated above uranium and thorium-rich deposits. The equivalent of the dose absorbed by its inhabitants amounts to 15 mSv per year.

Worked example 16.2.1

TREATING TUMOURS

A 10 g cancer tumour absorbs 2.5×10^{-3} J of energy from an applied radiation source. Calculate the dose equivalent if the source is an alpha emitter.

Thinking	Working
Convert the mass of the tumour from grams to kilograms.	$10\text{ g} = \frac{10}{1000} = 0.01\text{ kg}$
Calculate the absorbed dose using the formula: $\text{absorbed dose} = \frac{\text{energy absorbed}}{\text{mass of tissue}}$	$\text{absorbed dose} = \frac{\text{energy absorbed}}{\text{mass of tissue}}$ $= \frac{2.5 \times 10^{-3}}{0.010}$ $= 0.25\text{ Gy}$
Calculate the dose equivalent using the formula: $DE = AD \times QF$ $QF = 20$ for alpha particles	$DE = AD \times QF$ $= 0.25 \times 20$ $= 5.0\text{ Sv}$

Worked example: Try yourself 16.2.1

TREATING TUMOURS

A 25 g cancer tumour absorbs 5.0×10^{-3} J of energy from an applied radiation source. Calculate the dose equivalent if the source is an alpha emitter.

EFFECTIVE DOSE

The different organs of the body have different sensitivities to radiation doses. For example, if a person's lung was exposed to a dose of 10 mSv, it would be more than twice as likely that cancers could develop than if the same 10 mSv dose was delivered to the liver. The weightings assigned by the International Commission of Radiological Protection (ICRP) to the various organs are shown in Table 16.2.4.

Effective dose takes into account the sensitivity of the organ to ionising radiation. Effective dose (ED) is found by calculating the sum of the dose equivalents (DE) multiplied by the **weighting factor**, W, for each organ affected. In physics, the symbol sigma (Σ) is used to represent the 'sum of'.

$$\text{Effective dose} = \Sigma(\text{dose equivalent} \times W)$$

Effective dose is measured in sieverts (Sv)

Body part	Weighting (W)
ovaries/testes	0.20
bone marrow	0.12
colon	0.12
lung	0.12
stomach	0.12
bladder	0.05
breast	0.05
liver	0.05
oesophagus	0.05
thyroid	0.05
rest of body	0.07
total	1.00

TABLE 16.2.4 The ICRP weighing values.

PHYSICS IN ACTION

Iodine-131 and the thyroid

On 11 March 2011, a catastrophic earthquake and tsunami hit Japan, killing tens of thousands of people and severely damaging the nuclear power station at Fukushima. Radioactive materials, including caesium-137 and iodine-131, escaped into the surrounding environment. These have half-lives of 30 years and 8 days respectively.

Your body need iodine for the healthy functioning of the thyroid gland, which maintains proper metabolism. Foods rich in iodine include seafood, vegetables and salt. However, the body cannot tell the difference between normal iodine and radioactive iodine. To prevent the people in Japan from absorbing radioactive iodine into their thyroid glands, they were issued with iodine tablets. Taking an iodine tablet each day ensured that the thyroid gland was saturated with iodine and so any radioactive iodine ingested by eating contaminated food would not be taken into the body and deposited in the thyroid (see Figure 16.2.5).

Many victims of the Chernobyl nuclear disaster in 1986 died of thyroid cancer years after the accident. They ingested radioactive iodine and this accumulated in the thyroid gland, eventually leading to cancer. They had not been issued iodine tablets.

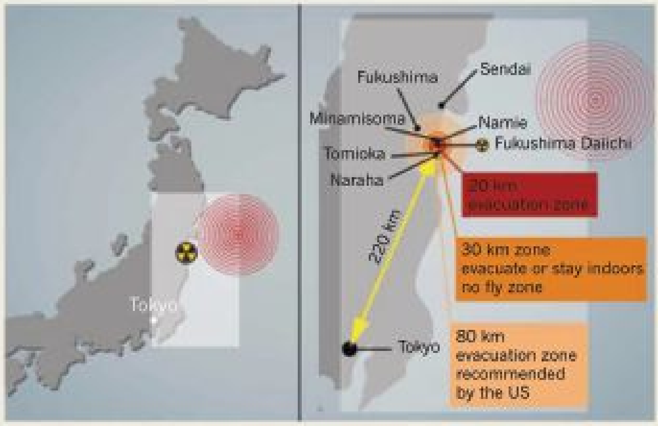


FIGURE 16.2.5 Children in Japan receiving iodine tablets after the Fukushima nuclear accident.

16.2 Review

SUMMARY

- Alpha, beta, gamma and high-energy electromagnetic radiation are ionising and are harmful to humans.
- Exposure to some ionising radiation is a natural part of human existence. However, unnecessary exposure to ionising radiation can be dangerous and should be avoided.
- The absorbed dose (AD) is a measure of the radiation dose per kilogram of irradiated tissue: Absorbed dose (AD) = $\frac{\text{energy absorbed by tissue}}{\text{mass of tissue}}$
- AD is measured in grays (Gy).
- The quality factor (QF) of radiation is a number that indicates the relative damaging effect of a particular radiation.
- The dose equivalent (DE) gives a measure of the biological damage that a dose of radiation causes. DE = AD × QF and is measured in sieverts (Sv).
- The effective dose (ED) is found by calculating the sum of the dose equivalents (DE) multiplied by the weighting factor, W, for each organ affected.
- Effective dose = Σ(dose equivalent × W) and is measured In sieverts (Sv)
- When ionising radiation passes through human tissue, it may ionise atoms and molecules in the body's cells, which can lead to the development of cancerous cells.
- Exposure to ionising radiation can lead to both somatic and genetic effects.
- Depending on the radiation dose, somatic effects can vary from feelings of nausea to severe illness and even death.
- If a person's reproductive cells are damaged by radiation, genetic abnormalities may arise in future generations.

KEY QUESTIONS

1 What is a somatic effect of exposure to ionising radiation? Give three examples.

2 A woman is exposed to a large whole-body dose of radiation and was later found to be anaemic (to have low red blood cell count). Is this a genetic or somatic effect? Explain your answer.

3 A cancer tumour of mass 150 g is exposed to 0.30 J of radiation energy. What is the absorbed dose in grays?

4 A geologist receives a dose of 250 μSv. Which of the following is the most damaging dose: 250 μSv of alpha, 250 μSv of beta, or 250 μSv of gamma? Explain your answer.

5 An 80 kg tourist absorbs a gamma radiation dose of 200 μGy on a return flight from London.

- **a** Calculate the dose equivalent received by the tourist.
- **b** Calculate the amount of radiation energy that has been absorbed by the tourist.

6 Which of the following is the most damaging radiation dose: 200 μGy of gamma radiation, 20 μGy of alpha radiation, 50 μGy of beta radiation or 30 μGy of neutron radiation? Explain your answer.

7 When in space, astronauts receive a radiation dose of about 1000 μSv per day. The normal annual background dose on Earth is 2 mSv.

- **a** How many days does it take for astronauts to exceed the normal annual background dose?
- **b** The record for time spent in space is 879 days held by cosmonaut Gennady Padalka. How much radiation in millisieverts (mSv) was he exposed to in this time?

8 To treat cancer of the uterus, a radioactive source is implanted directly into the affected region. If the uterus receives a dose of 0.40 Gy h^{-1} from the source, how many hours should it be left there to deliver a dose of 36 Gy?

9 Refer to Table 16.2.4 of W values. Which of these organs are most sensitive to ionising radiation: lung, bone marrow, bladder, ovaries? Explain your answer.

10 Refer to Table 16.2.4 of W values. Calculate the effective dose for a patient whose organs received the following exposures during a course of radiotherapy: ovaries and bladder 35 mSv each, colon 50 mSv.

16.3 Radiation in diagnosis and treatment of human disease

'Cancer' is a general term that actually incorporates different diseases. The term 'tumour' describes an abnormal mass of tissue due to the increased growth of cells. These can be either benign, containing only normal cells, or malignant, where growth of cancerous cells takes place.

Malignant tumours can grow in just about any part of the body and invade the surrounding healthy tissue. Moreover, some cancer cells may break away and be carried by the bloodstream; they then settle in other parts of the body, spreading the cancer (see Figure 16.3.1). It is well accepted that exposure to ionising radiation can cause cancer, but it can often also provide a cure.

Over the past 50 years, a variety of imaging techniques have been developed for diagnosing cancer, as well as several other diseases. These include computed tomography (CT), gamma (γ) radiation scans, magnetic resonance imaging (MRI), positron emission tomography (PET) scans and single photon emission computed tomography (SPECT).

Two forms of electromagnetic radiation, X-rays and γ rays, as well as alpha, beta and positron particles, are commonly used during the treatment of cancer. The application of these techniques in the diagnosis and treatment of disease is the focus of this section.

FIGURE 16.3.1 A bone scan of a patient with prostate cancer using technetium-99m. The skeletal metastases (cancers) take up the radioactive isotope differently to the rest of the body, and appear as black dots spread out across the entire skeleton: (a) anterior view; (b) posterior view.

X-RAY-BASED DIAGNOSIS AND TREATMENT

Diagnosis

In medical practice, soft X-rays are used for both diagnosis and treatment. Examples of medical radiographs that you may be familiar with are shown in Figure 16.3.2. It is possible to obtain these X-ray images due to the difference in the radiation absorption capacity of the tissues in your body. This is what creates an inverse shadow of your body when exposed to a radiographic film. Because of their thicker and harder consistency, bone and teeth are able to absorb and block X-rays when exposed to them. The surrounding softer tissues such as muscle, ligaments and skin do not block X-rays, thus the radiation travels straight through them.

FIGURE 16.3.2 Digital radiography systems are common nowadays.

During the 1970s, a new **diagnostic imaging** technique that also used X-rays was developed. It was called computed tomography (CT) scanning, also known as CAT (computerised axial tomography) scanning. The word *tomography* comes from the Greek term *tomos* meaning slice or knife. CT scans are an image of a slice of the body. To understand the usefulness of CT scans, imagine a large sausage like that shown in Figure 16.3.3. By taking a set of thin slices of the sausage you could see exactly what it was made of. If you saw anything unexpected in your slice (or cross-section), you could determine exactly where in the sausage it was originally located.

FIGURE 16.3.3 A slice of something provides lots of information about its contents.

The main principles behind obtaining a CT scan image include the following:

- A set of thin X-ray beams and detectors are used. The width of the X-ray beam determines the thickness of the 'slice' of the body that is imaged, typically about 3–5 mm.
- The X-ray source and detectors are located in a circle around the body.
- Beams passing through a cross-section of the body are detected by the array of detectors. The intensity of the X-ray beam that reaches each tiny detector is converted to digital information and is sent to a computer.
- The computer allocates a corresponding shade from a black–grey–white scale, according to the different radiation absorption capacity of the body.

The X-ray source and detectors are rotated through a small angle and the detectors record a set of readings. The computer uses a set of these readings to construct an image of the cross-section of the body for this narrow slice that has been analysed. An example of the image produced is shown in Figure 16.3.4. While this technique provides much more detailed images for doctors, the doses of radiation received by the patient are much higher than that for a standard X-ray radiograph.

Treatment

The X-rays used in therapy treatments are of a much higher frequency and energy than those used to produce diagnostic images, and have deeper penetrating abilities. In general, accelerating voltages of between 4 and 25 million volts (4–25 MV) are needed to produce these high-energy rays. This can be accomplished by using a linac machine (a linear accelerator). These linac machines rotate so that the beam of X-rays can be sent into a specific location inside the patient's body from different directions (see Figure 16.3.5). In this way, the treatment aims to deliver the maximum dose of radiation to the centre of the tumour.

FIGURE 16.3.4 A cross-sectional view from a CT scan.

FIGURE 16.3.5 A patient being prepared for a session of X-ray radiotherapy by a linac machine. The linac will move around as the dose is delivered.

RADIOPHARMACEUTICAL-BASED DIAGNOSIS AND TREATMENT

Diagnostic

In modern clinical practice radioisotopes, also called radionuclides, can be used to collect information about a person's physiological state. They can be used to investigate the functioning of specific organs, monitor certain biological processes (such as thyroid or liver function), or determine the advancement stage of a specific disease, such as cancer.

In order to achieve this, radioisotopes called **radioactive tracers** are attached to specific biomolecules or drugs, thus creating **radiopharmaceuticals**. These are then administered to a patient either orally or by injection (see Figure 16.3.6).

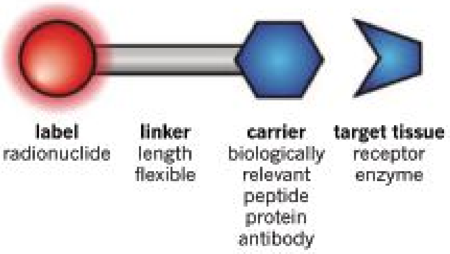


FIGURE 16.3.6 Illustration of a radiopharmaceutical, showing the radioactive label and the biomolecule or drug to which is attached.

The specific radioisotope used in the generation of a radiopharmaceutical depends on the tissue or organ that is suspected of malfunctioning or of growing abnormal cells. The body naturally distributes different molecules to different organs and this is used to target specific organs in the body. For example, iodine is normally sent to the thyroid gland by the liver. So if a radiopharmaceutical containing radioactive iodine is taken, most of this iodine will end up in the thyroid gland.

In order for a radioisotope to be used for diagnostic imaging it must:

- have a short half-life (hours or days) that is appropriate for the time taken for the diagnostic procedure. Radioactive materials are considered to be relatively safe after around 10 half-lives have passed
- emit only γ radiation of an energy that can be detected by a γ camera
- not emit alpha or beta radiation because these particles would be trapped in the patient's tissues and they would not be detected externally
- be available in the highest possible activity but not be toxic to the patient or react with drugs administered at the same time.

When the tracer has reached the target organ, a radiation scan is taken with a γ camera. An unusual pattern on the scan indicates a possible health problem.

Table 16.3.1 shows the most widely used radioisotopes in clinical practice and their main characteristics.

These radioisotopes can be then used in the following imaging techniques.

Medical radioisotope	Half-life	Radiation emitted	Medical use
molybdenum-99 (Mo-99)	66.02 hours	β	production of Tc-99
technetium-99m (Tc-99m)	6 hours	γ rays	widely used as a radioactive tracer for many organs
iodine-131 (I-131)	8 days	β and γ rays	thyroid-related illnesses
phosphorus-32 (P-32)	14 days	β	malignant tumour identification
chromium-51 (Cr-51)	28 days	γ rays	assessment of renal (kidney) function
gallium-67(Ga-67)	3.26 days	γ rays	Hodgkin's disease, lymphomas, lung cancer
arsenic-74 (As-74)	18 days	β^+	identification of brain tumours

TABLE 16.3.1 Characteristics and uses of radioisotopes.

PET scans

Positron emission tomography (PET) scanning is a medical imaging technique that gives an assessment of how an organ is working, rather than providing information about its structure. The technique is very sensitive for detecting the early stages of an illness and can detect malfunctions even without any structural changes in the tissue.

Emitted positrons travel for a short distance from their site of origin and lose energy while passing through tissue. When most of their kinetic energy has been lost, they undergo a process called annihilation. This happens when the positrons react with electrons from the immediate tissue area, and results in the emission of two very-high-energy photons in the form of gamma rays. These gamma rays are recorded by a γ camera and are reconstructed into images by a computer. The diagram in Figure 16.3.7 shows how this process works.

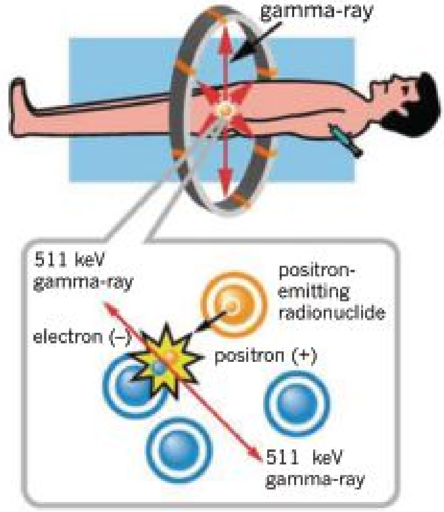
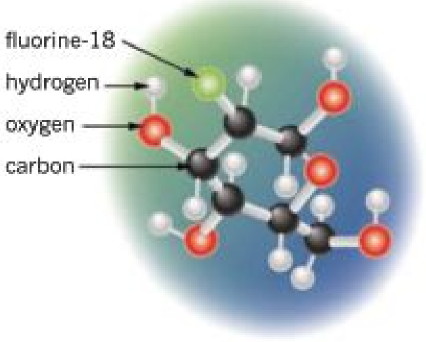


FIGURE 16.3.7 The basic principle behind PET scan technology.

The basic principle of PET scanning relies on the selection of a biologically relevant molecule which is selectively taken up by a specific organ or tissue under examination (for example, glucose, estradiol and methionine). Then, this molecule is labelled with a positron-emitting radiotracer, such as ^{18}F, ^{15}O, ^{82}Rb or ^{64}Cu.

For example, for PET scans of the brain, the radioactive tracer ^{18}F is attached to molecules of D-glucose, creating the radiopharmaceutical fludeoxyglucose (^{18}F-FDG), which is shown in Figure 16.3.8. This arrangement is chosen because the brain uses high levels of glucose in its metabolism.



2-fluoro-2-deoxy-D-glucose (FDG)

FIGURE 16.3.8 The ^{18}F-FDG radiopharmaceutical is the standard radiotracer used for PET imaging of the brain.

Other applications of brain imaging with 18F-FDG include that shown in Figure 16.3.9, where an FDG-PET scan is used to predict learning and memory problems that may arise with Alzheimer's disease.

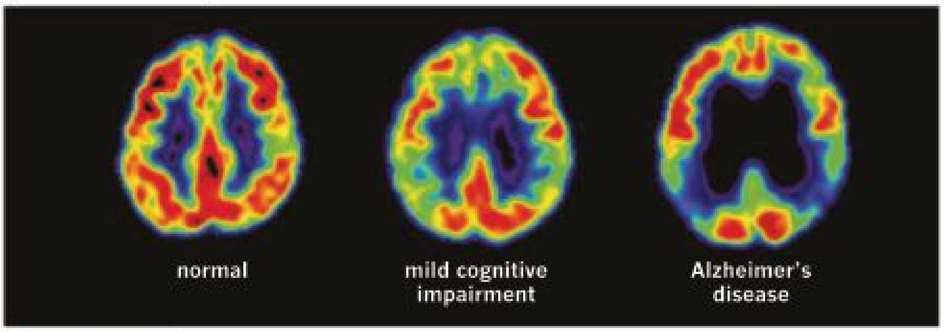


FIGURE 16.3.9 FDG-PET images showing reduced glucose metabolism in selected brain regions in patients with mild cognitive impairment and Alzheimer's disease. Areas with lowered uptake of FDG, and therefore reduced glucose metabolism, are shown in black in the images.

SPECT scans

A single photon emission computed tomography (SPECT) scan is a more-advanced type of nuclear imaging that is also used for diagnostic purposes. A radioactive tracer and a special camera are used to create 3D images of specific organs in the body. For example, SPECT can give information about blood flow to tissues and chemical reactions taking place in specific locations in the body (see Figure 16.3.10).

In a similar manner to PET scans, SPECT uses radioactive tracers attached to specific biomolecules and a scanner to record data and construct 3D images. Some examples of these images are shown in Figure 16.3.11. For the procedure to take place, a small amount of a γ emitter radiopharmaceutical with a short half-life (hours or days) is injected into a patient's vein. Shortly after, a γ camera is used to generate detailed images of areas inside the body where the radiopharmaceutical has been taken up by the cells.

Radioisotopes such as technetium-99m, fluorine-18 and iodine-123 are commonly used in SPECT scans. They have half-lives of 6 hours, 110 minutes and 2.8 days respectively.

FIGURE 16.3.10 A SPECT machine is used to take a brain scan of this child at the Royal Children's Hospital in Melbourne.

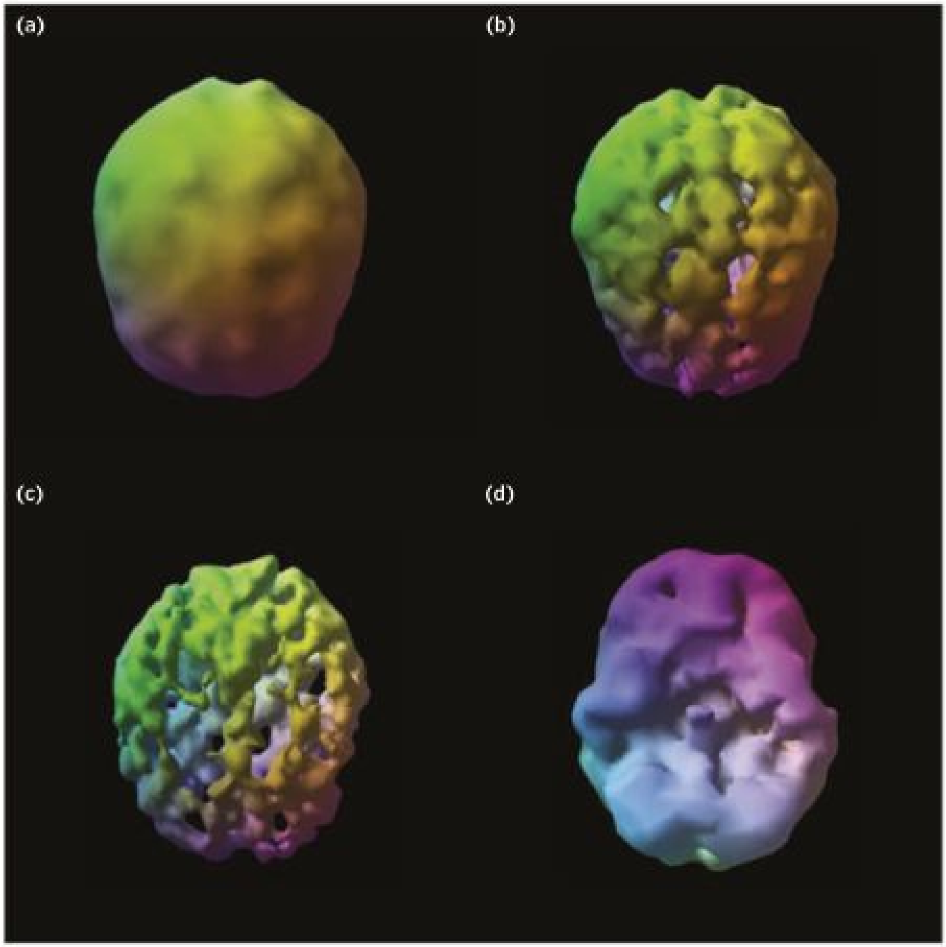
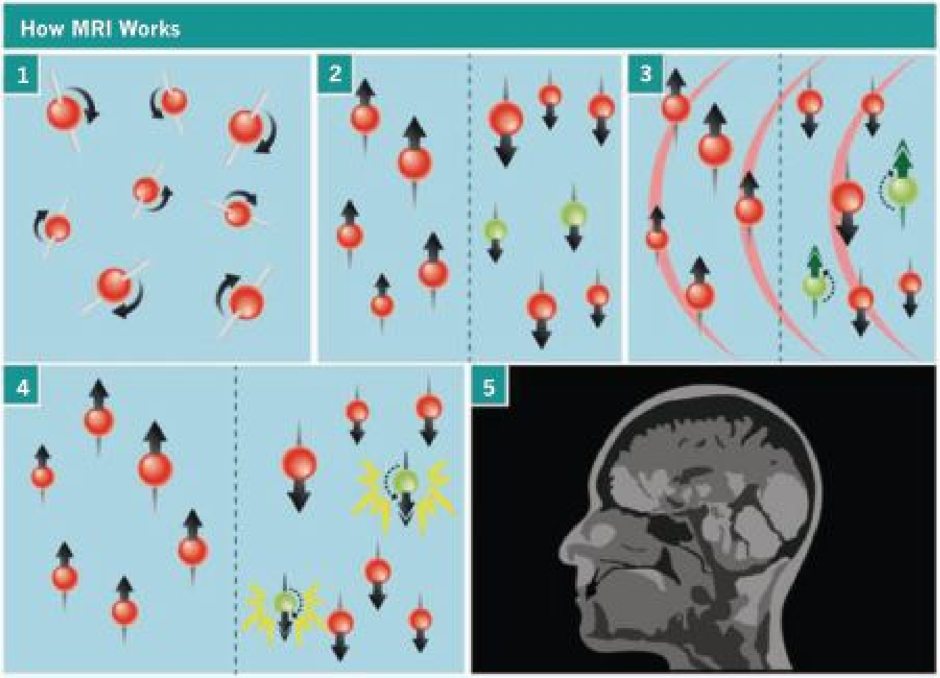


FIGURE 16.3.11 Brain SPECT scans showing blood flow and active neurological areas in brains under different conditions: (a) a healthy brain, showing normal blood flow and high neuronal activity in all regions of the organ; (b) a heavy methamphetamine abuser's brain; (c) a cocaine abuser's brain; and (d) a heavy smoker's brain. In all images the dents or holes correspond to brain areas with low metabolic activity, decreased blood flow and therefore impaired neuronal activity.

MRI scans

Magnetic resonance imaging (MRI) is a medical imaging technique used to investigate the anatomy of the body. It uses a magnetic field and radio waves to create 2D images of organs and tissues. It uses non-ionising radiation (radio frequency) and is therefore safer than X-ray imaging. It can be carried out using elements that have an odd number of nucleons, such as hydrogen, fluorine-19, sodium-23 and carbon-13. Since hydrogen is abundant in the body, it is used for most medical MRIs.

If a person is placed in a very strong magnetic field, certain protons in their body tissues will align with the magnetic field, a little like the way the needle of a compass aligns itself with a magnetic field. If the patient is then exposed to a pulse of low-energy electromagnetic radiation (radio frequency), the protons shift their alignment. After the pulse has passed, the protons move back to their magnetic field alignment. As they do so they emit low-frequency electromagnetic radiation, which is detected by the MRI machine. This emission is used to create images of the locations of the body tissue. The different tissues in the body have different concentrations of hydrogen; therefore, different body tissues will create contrasting shades on a magnetic resonance image. Figure 16.3.12 shows the process involved in producing an MRI image.



1. Atoms spin in random directions, like tops, around their individual magnetic fields.
2. In the magnetic field produced by MRI, the atoms line up either north or south. About half the atoms go each way, but there are a few unmatched atoms.
3. When a radio frequency pulse is applied, the unmatched atoms spin the other way.
4. When the radio frequency is turned off, the extra atoms return to their normal position, emitting energy.
5. The energy sends a signal to a computer. The computer uses a mathematical formula to convert the signal into an image.

FIGURE 16.3.12 The basic principle behind MRI technology.

Like CT scans, magnetic resonance images are cross-sections or slices of parts of the body. MRI scanners can take unusual oblique cross-sections of the body. Since they can discriminate very well between the different types of soft tissue, they are used for imaging different organs throughout the body (see Figure 16.3.13).

FIGURE 16.3.13 MRI is often used to investigate sports injuries because every type of body tissue can be clearly imaged, not just hard or soft tissue. This image shows an MRI scan of a healthy upper back.

Treatment

There are a variety of ways in which ionising radiation from radioisotopes can be used to treat diseases, especially cancer. The radiation can be applied internally or externally to a patient's body.

Applied externally

(a) Cobalt-60 external beam therapy

Cobalt-60 machines were first developed in the 1950s. A beam of gamma rays from a cobalt-60 source is directed from an external source through a patient into the tumour site.

- Cobalt-60 has a half-life of 5.3 years so the source lasts for a long time before it needs replacing.
- The source cannot be switched off, and it needs to be shielded using lead so that the gamma rays do not irradiate everyone in the region.

Nowadays, cobalt-60 machines have been replaced by linacs in many countries, including Australia.

(b) The Gamma Knife

Despite its name, the Gamma Knife isn't a real knife but a machine that delivers a focused, high dose of gamma radiation to its target, made possible by the 200 or so cobalt-60 sources contained in it. The overlapping beams destroy the tumour while the individual beams do minimal damage to healthy tissue. Figure 16.3.14(a) shows how the individual beams can be focused on the target site.

The first Gamma Knife in Australia was installed in 2010 at the Macquarie University Hospital in Sydney. These are commonly used as an external source for treating brain tumours (see Figure 16.3.14(b)).

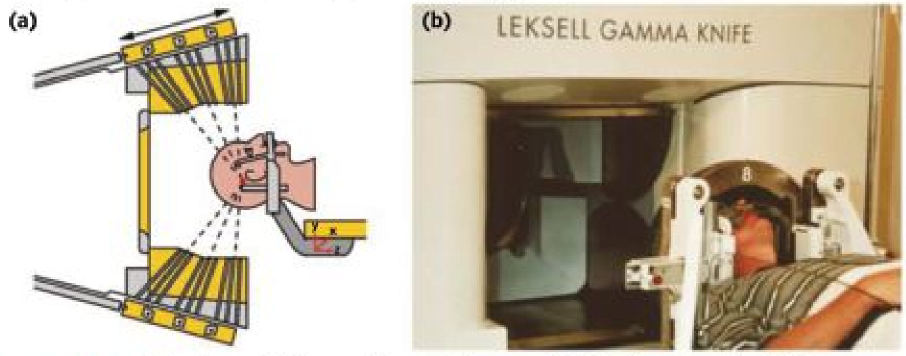


FIGURE 16.3.14 This Gamma Knife machine contains around 200 radioactive cobalt-60 sources and these beams overlap at the tumour site. (a) A schematic representation of how a Gamma Knife works. (b) A photograph of a patient undergoing a Gamma Knife procedure.

Applied internally

Radiopharmaceuticals may also be used in **chemotherapy**, where the radiation acts internally. In this case, the half-life of the radioisotope is an important consideration. Recall that radioactive materials take around 10 half-lives to become effectively non-radioactive. Therefore, a half-life of 5 seconds, for example, would not be suitable because the radioisotope would have almost completely decayed in less than a minute and not have a chance to do its job. A half-life of 5 years, on the other hand, would mean that the patient would be continually exposed to radiation for decades, which could cause more harm than good.

For a radioisotope to be used for this purpose, it must:

- have a short half-life (hours or days) that is appropriate for the time taken for the therapeutic procedure
- emit alpha or beta radiation, because these particles would be trapped in the patient's tissues and would destroy the cells in the tumour
- not emit too much gamma radiation because its high penetrating ability would mean that healthy cells and bystanders would be irradiated; however some gamma radiation allows a gamma camera to monitor the tumour.

Brachytherapy

Brachytherapy, also known as Radioactive Seed Implantation Therapy, is an advanced form of cancer treatment. It uses small wires or seeds made of radioactive elements that are temporarily or permanently implanted in the body at a site near or within the tumour growth area. From there they radiate ionising energy that kills the abnormal cells. Dozens of these seeds can be surgically placed around the tumour site, as shown in Figure 16.3.15. This method is most commonly used for tumours in the breast and prostate gland. There are two modes of delivery of this treatment:

- Low dose rate brachytherapy (LDB): Seeds are applied permanently and emit a low dose of radiation.
- High dose rate brachytherapy (HDB): Seeds are applied temporarily with high radiation potency.

Some of the clinically used radioisotopes now used in brachytherapy include iridium-192, iodine-125, cobalt-60 and cesium-137.

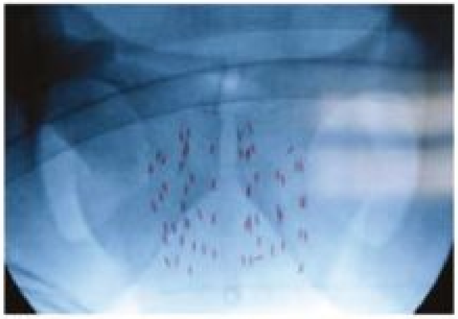


FIGURE 16.3.15 Numerous iodine-125 (I-125) wire seeds can be seen in this X-ray of a prostate cancer patient. I-125 is a gamma emitter with a half-life of 59.4 days.

16.3 Review

SUMMARY

- X-rays can be used for both diagnosis and treatment of disease. For diagnosis, X-ray radiographs and CT scans are routinely used, which employ soft X-rays. For the destruction of abnormal cells, beams of higher energy X-rays are delivered through linac machines.
- A radioisotope can also be called a radionuclide or radioactive tracer. All terms refer to a chemical element that has an unstable nucleus and emits radioactivity during its decay to form a stable element.
- A radiopharmaceutical refers to a drug that has a radioisotope covalently attached to its structure. The type of radioisotope used depends on the target tissue and the specific application. These drugs can be used for the diagnosis and treatment of disease.
- PET scans provide a metabolic assessment of specific organs in the body. The radionuclide used is a positron emitting molecule which undergoes annihilation when in contact with the target tissue. This creates two photons in the form of γ rays, which can be detected by a γ camera.
- SPECT scans use a radioactive tracer and a special camera and imaging system that creates a 3D image of specific organs in the body. It is used to give information about how organs work inside the body.
- MRI scans are used to investigate the anatomy of the body using two-dimensional images. It uses low-energy electromagnetic radiation (radio frequency) to shift the position of protons in tissues, which creates electromagnetic radiation that is detected by an MRI machine.
- Cobalt-60 machines and gamma knives are used to deliver high doses of γ radiation to a focused region of the body from the outside.
- Radiopharmaceuticals can also be used as part of chemotherapy treatments, where they deliver radiation from the inside of the body to kill the cancer cells on-site.
- Brachytherapy or Radioactive Seed Implantation Therapy is an advanced form of cancer treatment in which small wires or seeds made of radioactive elements are placed in very close to or inside a tumour. Depending on the dose needed, the seeds can be implanted temporarily (high-dose therapy) or permanently (low-dose therapy).

KEY QUESTIONS

1. How can an X-ray beam be used to treat a cancerous tumour and avoid exposing healthy tissue to a large dose of radiation?
2. Which type of electromagnetic radiation is used for a CT scan?
3. What are two characteristics that a radioisotope should have for it to be suitable for use as a tracer?
4. In which of the following nuclear medicine diagnostic techniques—X-rays, CT, PET, SPECT or MRI—would the medical workers need to take the lowest level of protection from harmful radiation?
5. What is a 'radiopharmaceutical'?
6. What type of radiation is emitted from the radioisotopes used in PET scans? Why is this?
7. Most forms of diagnostic imaging involve the patient being exposed to ionising radiation. If ionising radiation is dangerous, why is it used in diagnostic procedures?
8. Use Table 16.3.1 to determine which of the following radionuclides cannot be detected by a gamma camera: iodine-131, technetium-99, phosphorus-32.
9. How is an MRI scan similar to a CT scan? How are they different?
10. What is the radiation dose (if any) from an MRI scan?

Chapter review

16

KEY TERMS

absorbed dose
artificial transmutation
chemotherapy
cyclotron
diagnostic imaging
DNA
dose equivalent
effective dose
genetic
hard X-rays
linac
mutation
neutron bombardment
non-ionising radiation
particle radiation
photons
quality factor
radioactive tracer
radiopharmaceutical
soft X-rays
somatic
weighting factor

1 Classify alpha, beta, gamma, UV, cosmic rays and X-rays as either electromagnetic or particle radiation.

2 Compare the energy, penetrating power and speed of hard and soft X-rays.

3 Phosphorus-32 (P-32) is a radioisotope that is used for medical diagnosis of cancer. It is produced in the core of a nuclear reactor when a parent element X is bombarded with high-energy neutrons. What is the parent atom, X?

4 Are the γ rays emitted from a radioactive cobalt-60 source ionising or non-ionising and are they particle radiation or electromagnetic radiation?

5 Which form of radiation is able to alter the structure of atoms?

6 During a course of radiotherapy, a cancer patient may be exposed to 40 Sv of radiation. Why is this massive dose not fatal for the patient?

7 Which of the following is the most damaging dose of radiation: 1 Gy of alpha, 1 Gy of beta or 1 Gy of gamma? Explain your answer.

8 An oncology patient receives a radiation dose of 1 Sv. Which of the following is the most damaging dose: 1 Sv of alpha, 1 Sv of beta, or 1 Sv of gamma? Explain your answer.

9 An airline pilot of mass 90 kg absorbs a gamma radiation dose of 300 mGy during a return flight to New York. Calculate the dose equivalent that has been received in mSv.

10 In a major incident in a nuclear reactor, a 75 kg employee received a full-body absorbed radiation dose of 5.0 Gy. The radiation was gamma rays.
  a Calculate the amount of energy that was absorbed during this exposure. Give your answer to two significant figures.
  b Calculate the dose equivalent for this person. Give your answer to two significant figures.

11 A worker in an X-ray clinic takes an average of 10 X-ray photographs each working day and receives an annual radiation dose equivalent of 7900 μSv.
  a Calculate the dose (to one significant figure), in μSv, that the worker receives from each X-ray photograph. (Assume they work for 5 days per week for 45 weeks a year.)
  b How many times greater than the normal background radiation dose is the worker's annual dose?

12 Calculate the dose equivalent (in μSv) from a radiation source if the absorbed dose is 0.50 μGy and:
  a the radiation used is alpha radiation
  b the radiation used is beta radiation
  c the radiation used is gamma radiation.

13 During therapy for cancer, a patient's lungs receive 2500 μSv and his thyroid gland receives 1000 μSv. Use Table 16.2.4 on page 566 to calculate the effective dose of radiation for this patient.

14 What is the most widely used radioactive medical tracer and how is it produced?

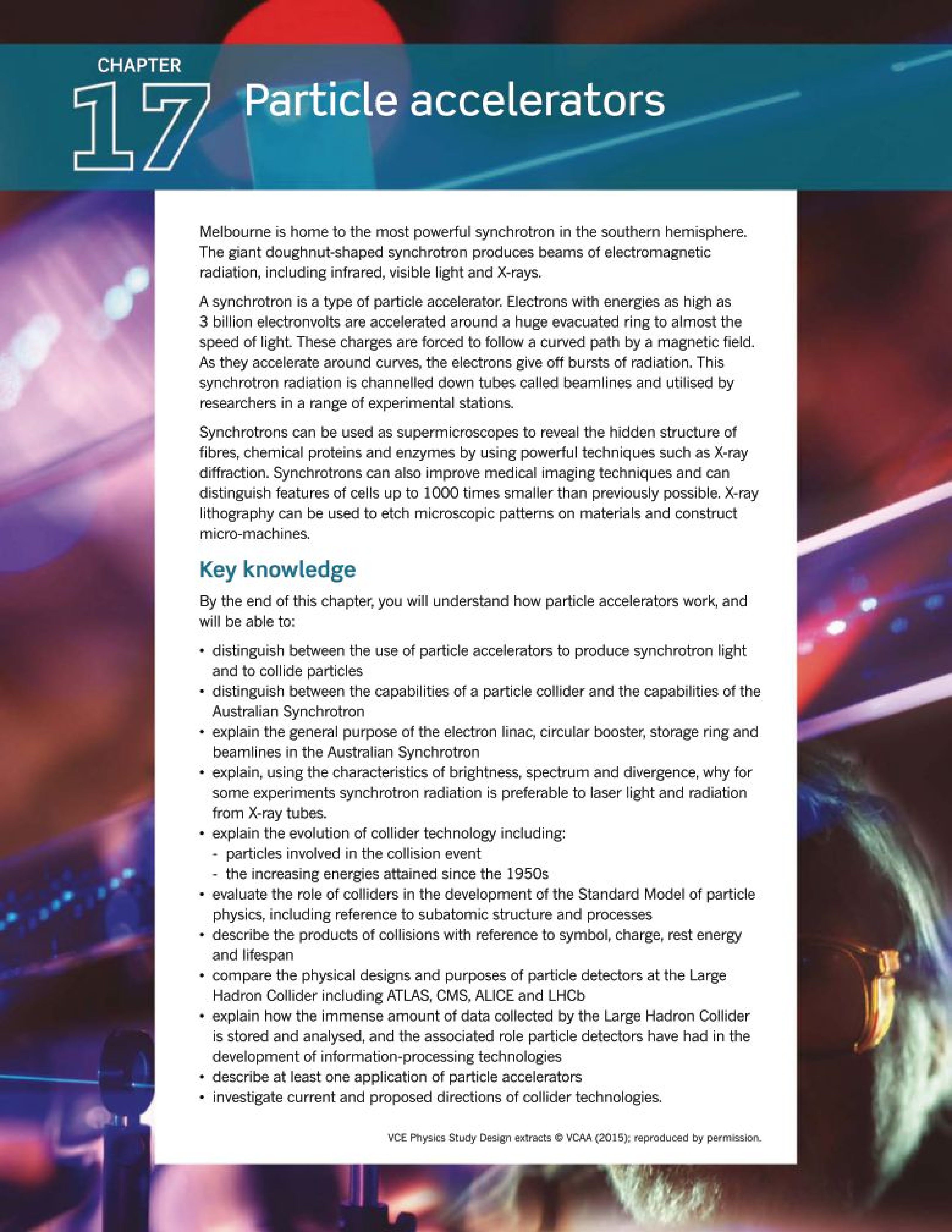
15 Do MRI scans use ionising or non-ionising radiation? Explain your answer by specifying the type of radiation used.

16 A radioisotope can be used as a tracer in the diagnosis of certain conditions in the human body. For this particular use, the radioisotope should ideally be a gamma emitter. Why is this?

17 What will a patient be injected with prior to undergoing a PET scan?

18 How do positrons interact with a patient's body in order to produce an image during a PET scan?

19 What are SPECT scans used for and how do they produce an image?



CHAPTER 17

Particle accelerators

Melbourne is home to the most powerful synchrotron in the southern hemisphere. The giant doughnut-shaped synchrotron produces beams of electromagnetic radiation, including infrared, visible light and X-rays.

A synchrotron is a type of particle accelerator. Electrons with energies as high as 3 billion electronvolts are accelerated around a huge evacuated ring to almost the speed of light. These charges are forced to follow a curved path by a magnetic field. As they accelerate around curves, the electrons give off bursts of radiation. This synchrotron radiation is channelled down tubes called beamlines and utilised by researchers in a range of experimental stations.

Synchrotrons can be used as supermicroscopes to reveal the hidden structure of fibres, chemical proteins and enzymes by using powerful techniques such as X-ray diffraction. Synchrotrons can also improve medical imaging techniques and can distinguish features of cells up to 1000 times smaller than previously possible. X-ray lithography can be used to etch microscopic patterns on materials and construct micro-machines.

Key knowledge

By the end of this chapter, you will understand how particle accelerators work, and will be able to:

- distinguish between the use of particle accelerators to produce synchrotron light and to collide particles
- distinguish between the capabilities of a particle collider and the capabilities of the Australian Synchrotron
- explain the general purpose of the electron linac, circular booster, storage ring and beamlines in the Australian Synchrotron
- explain, using the characteristics of brightness, spectrum and divergence, why for some experiments synchrotron radiation is preferable to laser light and radiation from X-ray tubes.
- explain the evolution of collider technology including:
  - particles involved in the collision event
  - the increasing energies attained since the 1950s
- evaluate the role of colliders in the development of the Standard Model of particle physics, including reference to subatomic structure and processes
- describe the products of collisions with reference to symbol, charge, rest energy and lifespan
- compare the physical designs and purposes of particle detectors at the Large Hadron Collider including ATLAS, CMS, ALICE and LHCb
- explain how the immense amount of data collected by the Large Hadron Collider is stored and analysed, and the associated role particle detectors have had in the development of information-processing technologies
- describe at least one application of particle accelerators
- investigate current and proposed directions of collider technologies.

17.1 Synchrotrons

A **particle accelerator** is a device that uses electromagnetic fields to propel charged particles to extremely high speeds, often close to the speed of light. Large accelerators are best known for their use in particle physics as colliders, including the highly publicised investigation that led the way to the discovery of the so-called 'god particle', or **Higgs boson**. In contrast to these types of particle accelerators that were built for collisions, a synchrotron light source is designed to use electrons to generate beams of infrared, UV, visible and X-ray radiation. It is this radiation, called synchrotron light, that is then channelled off for use in experimental stations to analyse materials.

Synchrotrons were first mentioned in Chapter 7 'Energy from the atom'. In this section, the key features and operation of a synchrotron are explored in more detail.

CATHODE RAY TUBES (CRT)

The **cathode ray tube** (CRT) of an old-style computer screen or television is a relatively simple type of particle accelerator. It can be used to explain the basic theory behind how larger and more advanced particle accelerators work.

All particle accelerators require a source of charged particles. In a cathode ray tube these are provided by a device called an electron gun. Electrons are 'boiled' off a heated wire element acting as the cathode (the negatively-charged electrode). The electrons are then accelerated from rest across a chamber emptied of air towards a positively-charged plate acting as the **anode** (the positively-charged electrode) by an electric field created between the charged plates.

As seen in Figure 17.1.1, the forces acting on the beam of charged particles between the plates will change the direction of the beam. In addition, as the charged particles are accelerated towards the top plate, work is done and the charges gain kinetic energy. By varying the electric field between the plates, by changing the charge on these plates or by adding electric or magnetic fields around the beam of electrons, the direction and width of the electron beam can be further controlled.

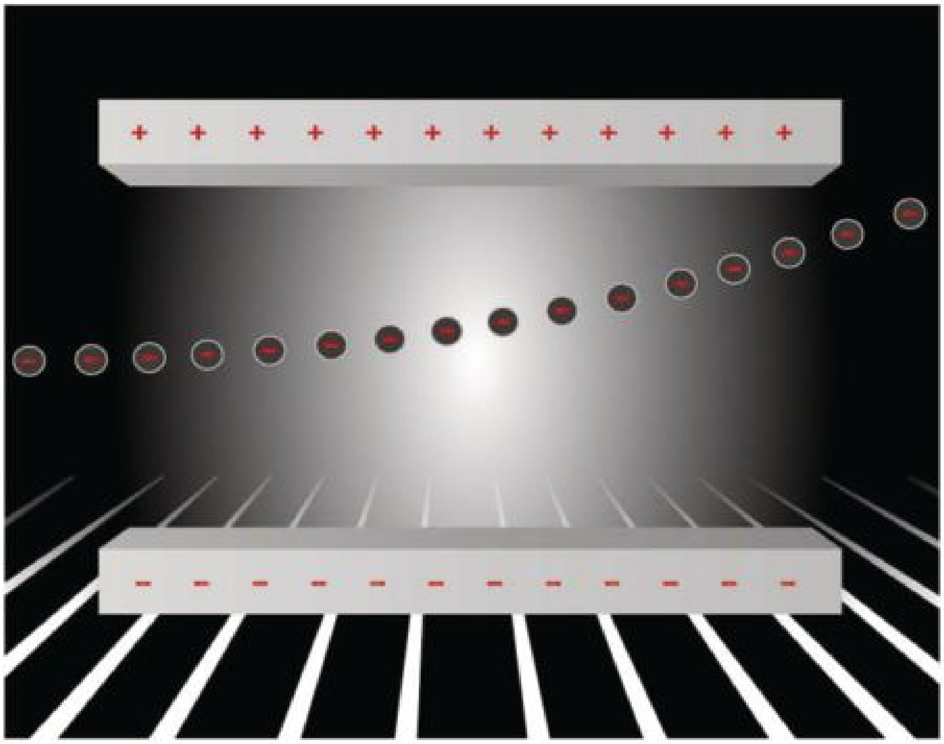


FIGURE 17.1.1 This beam of negatively charged electrons enters from the left into the electric field between two charged plates. It follows a curved path as it is repelled or attracted by the charge on the plates. It moves at a constant horizontal velocity but is accelerating towards the plate with the opposite charge to itself.

SYNCHROTRONS

Synchrotron light was first discovered in the 1940s, when it was observed being produced in particle accelerators used for theoretical physics. When first discovered, this radiation was seen as an unwanted by-product of the acceleration process, as its release robbed accelerating particles of energy. It was only later that the useful benefits of such radiation became apparent. Since their origins in the 1940s, synchrotrons have undergone progressive evolution.

Originally, large particle accelerators were designed to investigate the nature of matter by examining the structure of atoms and molecules via collisions. Strong magnets directed the particles to collide with a target or with another moving particle. Scientists used the data gained from these collisions to learn more about the make-up of the subatomic particles fired from the machine or the target samples that were hit.

In contrast to these types of particle accelerators built for collisions, a synchrotron light source is designed to use electrons to generate beams of infrared, UV, visible and X-ray radiation. Synchrotron light, or synchrotron radiation, is the term given to a range of electromagnetic radiation of wavelengths from approximately 10^{-3} to 10^{-10} m. This electromagnetic radiation is produced by charged particles such as electrons or protons as they travel in a curved path at speeds close to that of light. The beam of the synchrotron light produced falls in the shape of a cone ahead of the travelling charged particles.

The Australian Synchrotron at Monash University, shown in Figure 17.1.2, is an example of this type of particle accelerator.

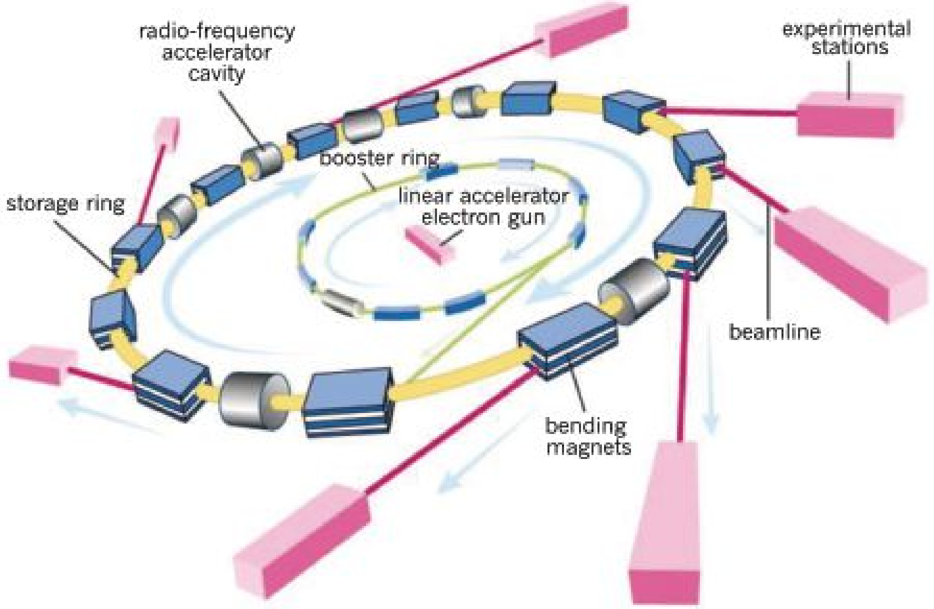


FIGURE 17.1.2 The Australian Synchrotron is about the size of a football field. This scale is necessary to contain the electrons, which are travelling at almost the speed of light as they zoom around the storage ring. This diagram shows the main features and experimental stations arranged around the storage ring.

The electron linac

Cathode ray tubes are useful particle accelerators but are limited to using voltages over a few tens of kilovolts. A linear accelerator, or **linac**, accelerates particles in straight lines. The first linac was built in 1928 by the Norwegian engineer Rolf Wideröe (Figure 17.1.3). It consisted of three hollow metal tubes inside an evacuated cylinder. These are called drift tubes and were used in Wideröe's machine to accelerate potassium ions to an energy of 50 000 eV (50 keV). This type of accelerator is referred to as a standing wave linac.

FIGURE 17.1.3 Rolf Wideröe, 1902–96.

PHYSICSFILE

The electronvolt

The unit **electronvolt** (symbol eV) is regularly used to describe the energy gained by a charged particle in a particle accelerator. It shouldn't be confused with the unit 'volt' familiar from the electrical circuit section of this course. An electronvolt is a unit of energy equal to the energy acquired by an electron moving through a potential difference of one volt. Since $\Delta E_k = W = qV$ and the charge on an electron is approximately 1.6×10^{-19} C, then 1 eV is:

$$\begin{aligned}\Delta E_k &= W = qV \\ &= 1.6 \times 10^{-19}\ \text{J}\end{aligned}$$

So, 1 electronvolt is a very, very small amount of energy.

The Australian Synchrotron and other electron linacs make use of travelling waves rather than standing waves in order to accelerate particles. The travelling-wave linac consists of:

- an electron gun
- a vacuum system
- focusing elements
- RF (radio-frequency) cavities.

As shown in Figure 17.1.4, electrons escape from the electron gun as they boil off the heated filament of the assembly. From here, they accelerate across a potential difference of about 100 keV and exit the electron gun at a velocity of approximately half the speed of light. At such velocities, the effects of relativity come into play. (The effects of relativity are outside the scope of this course but will be described in more detail in Units 3 and 4.)

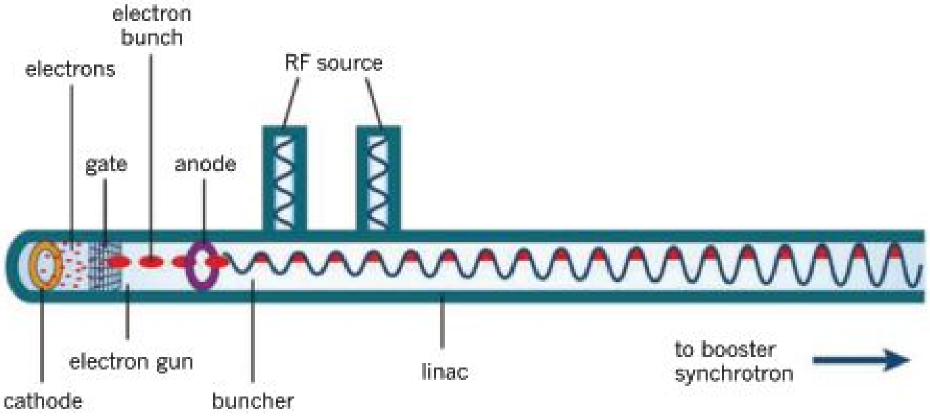


FIGURE 17.1.4 A diagram of a travelling-wave linac.

The electron beam travels through an ultra-high vacuum within the linac, to prevent energy loss through interaction with air particles. As the electrons travel, focusing elements act on the beam to constrict it to a narrow beam in the centre of the vacuum tube.

Electrons are accelerated to close to the speed of light by the end of the linac. Such acceleration is critical to the production of synchrotron light in the storage ring. This huge acceleration is achieved by cylindrical RF (radio-frequency) cavities that surround the electron beam. These cavities produce intense electromagnetic radiation at several hundred megahertz (MHz). The RF radiation propagates through the linac as a travelling wave. When timed correctly, electrons can, in effect, 'ride the crest' of this RF wave, resulting in their acceleration to enormous speeds.

In the Australian Synchrotron, electrons are released from the gate of the electron gun in pulses every 2×10^{-9} s to travel towards the anode. These electrons accelerate as they pass through the RF cavity with the crest of the RF radiation and are slowed down when they pass through with the trough of the RF radiation. This effect causes the electrons to become bunched into groups as they travel through the linac itself. The frequency of RF radiation is timed to accelerate the arrival of each electron bunch. The linac used in the Australian Synchrotron gives the electrons a kinetic energy of 100 MeV.

The booster ring

Within the circular **booster ring** of the Australian Synchrotron, bending magnets provide a force at right angles to the motion of the electrons in order to bend them into a circular path. In this ring, the energy of the electrons is boosted from 100 to 3000 MeV, or 3 GeV (gigaelectronvolts). The energy boost is supplied by an RF chamber through which the electrons travel on each orbit of the ring.

The storage ring

The booster ring channels the electrons into the **storage ring**, a doughnut-shaped tube (see Figure 17.1.5). In the Australian Synchrotron this ring has a radius of 34.3 m and a circumference of 216 m. Around this ring are 14 bending magnets, each 1.7 m long. These keep the electrons in the circular path. They are separated by 14 straight sections in which focusing magnets keep the electrons confined to a flat beam less than half a millimetre wide and only two-hundredths of a millimetre high.

FIGURE 17.1.5 The booster to storage ring transfer line in the Australian Synchrotron.

In the storage ring of the synchrotron, electrons orbit for hours at a time at speeds near that of light. A series of magnets makes them bend in arcs as they travel through the ring. As the electrons change direction (i.e. accelerate), they emit synchrotron radiation.

Several different types of magnets are used to direct the beam of charged particles. Bending magnets, called dipole magnets, guide the particles through small arcs that combine to produce 360° of bending around the ring. Other magnets called quadrupole and sextupole magnets refocus the beam to prevent it diverging and keep the particles in stable orbits. Steering and corrector magnets correct and fine tune the orbit to one-millionth of a metre. The specific arrangement and strength of the magnets, called the lattice, dictates characteristics of the synchrotron light produced, including:

- brightness
- polarisation
- energy distribution
- coherence.

As the electron beam circulates, it radiates synchrotron light and loses energy. To counter this, the electrons pass through RF cavities, like those in the booster ring, to replenish the energy lost. The RF cavities have electromagnetic fields oscillating at radio frequencies produced by amplifiers located next to the storage ring. These fields oscillate in polarity extremely quickly, up to 500 million times per second (500 MHz). The process ensures that the electrons stay at a constant energy and remain stored in the ring.

Despite the RF cavities, the beam is still not perfectly stable. All synchrotron beams will gradually reduce in intensity with time. Some electrons are lost in collisions between electrons and gas molecules in the near vacuum of the ring. To minimise these losses, the vacuum chamber must be kept at a pressure of about one-thousandth of one-billionth of normal atmospheric pressure, or less than 10^{-7} Pa. Under these conditions, the beam typically loses half of its intensity over a 5–50 hour period. New electrons are injected into the beam at 4–24 hour intervals to replace those lost through collisions and energy losses.

The unused high-energy X-rays given off by the storage ring are continually absorbed by radiation shielding. The shield wall surrounding the storage ring is usually made of lead and concrete forming a tunnel that completely encloses the storage ring, except for the beamlines through which radiation is guided. This design feature is critical for employee safety during synchrotron operation.

Beamlines

A beamline is the path that synchrotron light travels from the storage ring, where it is produced, to its target experimental work (see Figure 17.1.6). The point at which the beamline meets the storage ring is called the front end. A beamline is typically a stainless steel tube, 15–35 m in length and around 4 cm in diameter. The dimensions depend greatly on the technique being performed on the beamline and the application of that technique. A typical beamline consists of an optics room, an experiment room and a control room.

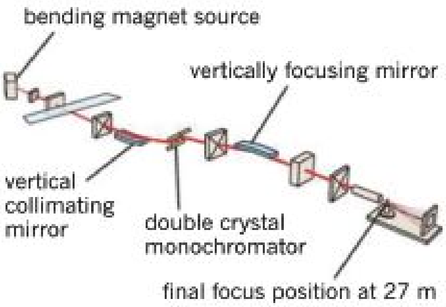


FIGURE 17.1.6 This diagram shows the arrangement of mirrors and crystal monochromator in the beamline for studying proteins and microcrystals and small molecule X-ray diffraction in the Australian Synchrotron.

Inside the optics room, synchrotron light is modified according to the needs of its experimental use. Sometimes scientists will wish to use only a specific range of wavelengths of synchrotron light for their experiments, rather than all of the light produced. A device called a monochromator, either a crystal or a grating, is used as a wavelength selector. As a beam hits this device, particular wavelengths are diffracted at different angles. By rotating the monochromator, a specific light frequency can be selected from the broad band of frequencies available in the incident beam. As it is prepared for its role in an experiment, synchrotron light may also be:

- aligned using slits
- refocused using mirrors
- lessened in intensity using attenuators within the optics room.

Thin beams of synchrotron light are directed onto a specific target or sample being examined within an experiment room. Scientists control their experiments from an external control room in which they are protected from the intense electromagnetic radiation being used in the experiments. All synchrotrons have a number of beamlines, each directing the synchrotron light to an experimental station. The Australian Synchrotron started out with 13 beamlines but has been designed to allow for additional beamlines as demand for experimental time increases.

CHARACTERISTICS OF SYNCHROTRON LIGHT

In Chapter 7 'Energy from the atom', synchrotron light was explained as the name given to electromagnetic radiation emitted when charged particles, such as electrons, are accelerated in curved paths. For high-energy electrons, the photons emitted have energies ranging from the infrared through to X-rays. Synchrotron light has a number of specific attributes that make it the preferred option for a range of experimental techniques:

- It has a broad spectral range, from infrared light to X-rays (see Figure 17.1.7).
- It has high intensity or brightness—hundreds of times brighter than from standard X-ray tubes (see Figure 17.1.8).
- It has a high degree of collimation and very low beam divergence.
- It is tunable, meaning specific frequencies can be selected from synchrotron light.
- It is emitted in very short pulses of less than a nanosecond.
- It is highly polarised—either linearly, circularly or elliptically.

The range of photon energies produced in the Australian Synchrotron has wavelengths corresponding to the dimensions of cells, viruses, proteins and atoms, enabling scientists to explore the structure of these objects. This is not possible with longer wavelength visible light. Short wavelength X-rays produced in synchrotron light are an ideal tool for examining structures at a cellular or atomic level. The brightness and monochromatic nature of synchrotron radiation make it ideal for investigating crystalline structures, using a technique called **X-ray diffraction**. The high intensity means that a particular analysis can be completed in far shorter time than with radiation from another source, such as an X-ray tube.

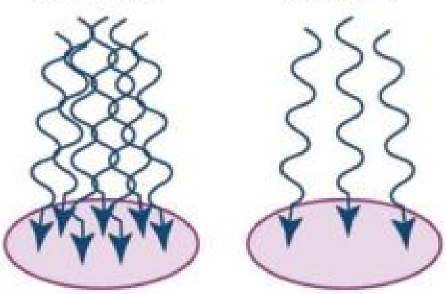


FIGURE 17.1.8 X-rays from a synchrotron source are far brighter than X-rays produced by an X-ray tube because a higher intensity is concentrated in a smaller area.

Synchrotrons also produce a continuous spectrum of radiation. This makes them more useful than lasers, which only produce very specific wavelengths. Different elements will absorb energy of a specific frequency. By being able to select particular wavelengths of synchrotron light, researchers can select the best wavelength or range of wavelengths for a specific technique or analysis.

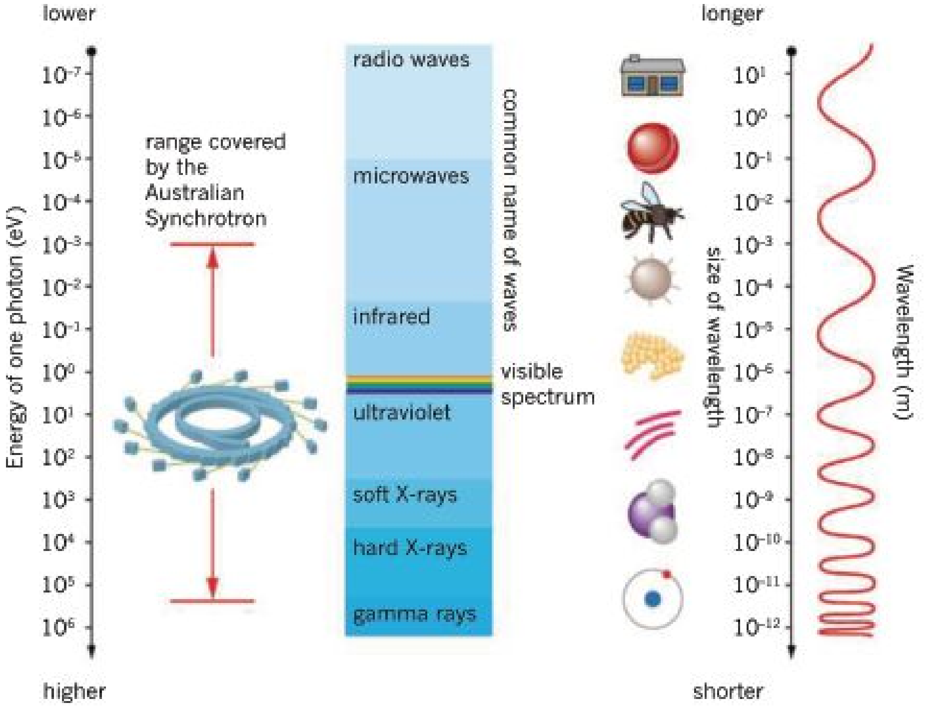


FIGURE 17.1.7 The range of wavelengths and photon energies produced as synchrotron radiation.

X-ray diffraction

One particular technique where synchrotron light has specific advantages is in the field of X-ray diffraction.

X-ray diffraction was initially developed by a father and son team from Adelaide, Sir William Henry and Sir William Lawrence Bragg (Figure 17.1.9). For their services in the analysis of crystal structure by means of X-rays, they were jointly awarded the Nobel Prize in Physics in 1915.

When a collimated X-ray beam falls on a single plane of atoms in a crystal, such as sodium chloride, each atom of the layer will scatter a small portion of the beam. The result is that the beam is scattered in many different directions. For most directions, the scattered light cancels out. However, at particular angles a reflected beam can be detected. At these angles, the scattered reflections from each atom are adding together, termed constructive **interference**, to produce a high-intensity reflected beam. By using X-rays of a known wavelength and rotating the crystal under investigation through all angles with respect to the incoming beam, a pattern of intensity peaks is produced, built up from X-rays reflecting from many layers within the crystal. The result is an X-ray diffraction pattern. The complete diffraction pattern generated in such a process is unique to the particular crystal and can be used to determine the 3D structure of the molecule (see Figure 17.1.10).

Synchrotron X-rays are the ideal tool for X-ray diffraction. A conventional X-ray is low intensity and not coherent, meaning that there is a spread of frequencies. By comparison, synchrotron X-rays are highly aligned, coherent and have a higher intensity. In addition, any desired wavelength within the synchrotron's spectrum can be specifically selected, whereas X-ray sources produce radiation at only a few specific wavelengths.

Synchrotron X-rays have suitable energies to interact with many common smaller atoms, like carbon and oxygen, whereas conventional X-rays have specific energies that will interact with heavier elements. Synchrotron X-rays are also around 100 million times brighter than those from conventional sources. The high brightness is due to the synchrotron X-rays being concentrated in a much smaller area. A synchrotron beam has greater brilliance than a traditional X-ray beam, meaning that it has a higher intensity per unit area.

FIGURE 17.1.9 (a) Sir William Henry, 1862–1942, and (b) Sir William Lawrence Bragg, 1890–1971.

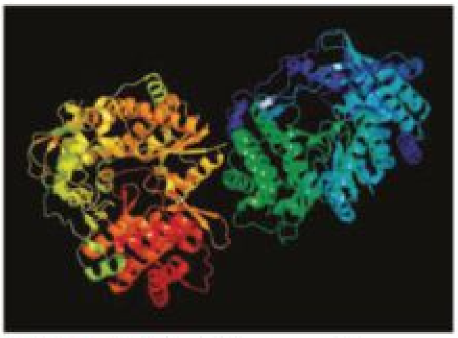


FIGURE 17.1.10 The 3D structure of the RNA-producing enzyme polymerase. The structure of this enzyme was determined by diffraction methods from a synchrotron source. Polymerase is responsible for replicating the gene of the hepatitis C virus. Now that its structure is known, scientists are trying to develop a drug that will inhibit its function and so stop the virus from replicating.

17.1 Review

SUMMARY

- All particle accelerators require a source of charged particles. In a cathode ray tube these are provided by a device called an electron gun.
- Force acting on the charged particles from electric and magnetic fields will change the direction of the beam. As the charged particles are accelerated by these forces, work is done and the charges gain kinetic energy.
- Originally, large particle accelerators, termed colliders, were designed to investigate the nature of matter by examining the structure of atoms and molecules via collisions. In contrast, a synchrotron light source is designed to use electrons to generate beams of infrared, UV, visible and X-ray radiation.
- A synchrotron is a doughnut-shaped particle accelerator designed to circulate electrons around a closed path at speeds very close to that of light.
- Electrons are emitted from an electron gun in pulses. They are accelerated through the linac by powerful bursts of radio-frequency (RF) radiation.
- Electrons then travel around a booster ring, and are accelerated further as they pass through RF cavities until reaching an energy of 3 GeV. The magnetic field of the bending magnets is increased as the velocity of electrons increases within the booster ring.
- Electrons are channelled into the storage ring. Synchrotron radiation (light) is produced as the particles travel through the strong magnetic fields of the dipole magnets.
- Electrons replace energy lost due to the production of synchrotron light as they pass through RF cavities in the storage ring.
- Synchrotron light leaves the storage ring, passing down beamlines to a number of independent experimental stations.
- The speed of the accelerated electrons can be found by equating the energy in eV with the kinetic energy gained by the electrons, i.e. $E_k = \Delta E$ or $\frac{1}{2}mv^2 = eV$.
- Synchrotrons are an ideal source of X-rays that can be used in a wide range of investigative applications. Compared with other sources, synchrotron light produces X-rays and other radiation that:
  - is of very high intensity and is produced in a beam that has a lifetime of up to several hours
  - has a frequency range from the infrared region of the spectrum through to X-rays
  - is highly collimated
  - travels in short pulses
  - has very high intensity
  - can be selected for specific frequencies and is highly polarised.

KEY QUESTIONS

1 What type of particles do all particle accelerators use?

2 In a cathode ray tube, how are particles accelerated?

3 List the four main features of the Australian Synchrotron in order from the initial emission of electrons at the cathode through to the experiment and control rooms.

4 In the booster ring, to what kinds of speeds can the electrons be accelerated to?

5 Describe the role of the storage ring.

6 In which section of the Australian Synchrotron is the strength of the magnetic field increased as the velocity of the electrons increases?

7 In which section of the Australian Synchrotron are electrons accelerated by RF radiation?

8 Compared with X-rays from conventional sources, X-rays from synchrotron radiation have a considerably higher intensity. What benefit does this offer researchers?

9 What specific attributes of synchrotron radiation make synchrotron sources more useful than laser sources for particular investigations of elements that absorb particular wavelengths?

10 Copy the table and list the properties of X-rays emitted by a conventional X-ray tube and a synchrotron.

X-ray tube	Synchrotron

17.2 Colliders and particle physics

Since the announcement of the proof of the existence of the Higgs boson, colliders have been the superstars of particle accelerators. The best known of the colliders is the Large Hadron Collider (LHC). Figure 17.2.1 shows an image of particles that were created in the LHC. The LHC is only one of a group of accelerators housed at CERN, Geneva, near the French–Swiss border. Since 1934, colliders have been at the forefront of experimentation in particle physics. This research that has led not only to a better understanding of the fundamental nature of matter but also to a range of discoveries that are being applied in areas as diverse as medicine and environmental science.

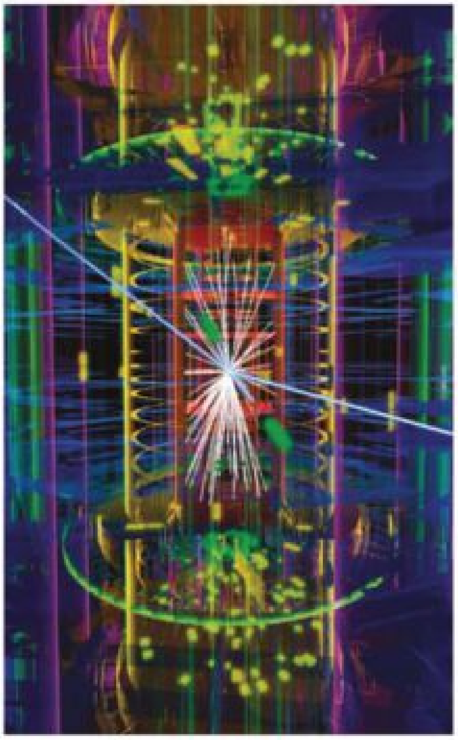


FIGURE 17.2.1 This image shows the shower of particles that were created following a particle collision in the ATLAS detector. The particles were accelerated in the LHC at CERN.

THE EVOLUTION OF COLLIDERS

Cyclotrons

The cathode ray tube (CRT) and the Van de Graff generator were arguably the first particle accelerators. The cathode ray tube was developed by German Physicist Ferdinand Braun in 1897. The Braun tube, as it was originally known, directed electrons from an unheated cathode in an evacuated glass tube, through a magnetic field onto a phosphor-covered screen. The magnetic field was used to change the direction of the electrons and accelerate them towards a particular portion of the screen. Varying the strength of the magnetic field changed the acceleration of the electrons. The basic principles of the Braun tube became the basis of cathode ray tube televisions and computer screens.

Braun's development of the CRT followed J. J. Thompson's discovery of electrons in the same year. To physicists, the discovery of the electron suggested that atoms could be broken down into smaller and smaller particles and the science of particle physics was born. During the 1920s, physicists believed that electrons and the earlier-discovered protons were the fundamental particles within an atom. However, discoveries in the 1930s and 1940s quickly made physicists rethink this idea, and the study of elementary particle physics began. Into the field of particle physics came the **cyclotron**, a device specifically developed to allow further investigation of fundamental particles.

An American physicist, Ernest Lawrence, was the first to produce a working cyclotron in 1932. This first cyclotron was just 13 cm in diameter and accelerated protons to 80 keV (1 keV = 1.6×10^{-16} J). Figure 17.2.2 gives a good comparison with modern-day particle accelerators.

Unlike Van de Graff generators that only accelerate electrons once via the voltage on the dome of the generator, the cyclotron accelerates the particles over and over, leading to particle energies many times that of the original. In Lawrence's cyclotron, protons were injected into the centre of a cylindrical space between the poles of a large electromagnet. The magnetic field caused the particles to move in a circular path, hence the name. An electric field accelerated the particles further, causing them to move in an outwards spiralling path towards the outside edge of the cyclotron, with ever increasing energy. At the rim, the particles were directed out of the cyclotron to hit a target. The high-energy particles colliding with the target generated secondary particles, which could then be studied.

It is hard to imagine that the start of developments in the study of particle physics in the Large Hadron Collider (LHC) came from an instrument just 13 cm across. Lawrence continued to develop successive models of the cyclotron, which increased the size and the energies produced. By 1945 he'd built a 4.7 m 'synchrocyclotron', which accelerated particles to energies of 730 MeV.

As the study of particle physics accelerated in the 1950s, so did the energies developed by successive generations of particle accelerators. To date, the largest cyclotron ever built, an 18 metre unit at the University of British Columbia in Canada, remains in operation and produces protons with energies of 500 MeV. Until the 1950s, when it was superseded by the synchrotron, the cyclotron was the most powerful particle accelerator available to particle physics.

FIGURE 17.2.2 In 1930, physicist Ernest Lawrence built the first successful cyclotron, at Berkeley, California. It was just 13 cm in diameter and it accelerated protons to an energy of 80 keV.

PHYSICS IN ACTION

Cyclotrons

To perform experiments at very high energies, linear accelerators would need to be extremely long. For this reason, in the 1930s the American physicist Ernest O. Lawrence designed the first circular accelerator. This is called a cyclotron, and it won Lawrence a Nobel Prize in 1939. In some respects, the cyclotron operates as a spiral-shaped linac. Protons are often used as the accelerating particles in this machine.

Here, the many drift tubes are replaced by two semicircular, D-shaped, hollow copper chambers, called 'dees'. These are the positive and negative electrodes of the cyclotron between which exists a strong electric field. The dees sit back to back, giving the cyclotron its circular shape, and lie between the poles of a powerful electromagnet. The inside of the metallic dee is shielded from the electric field. The magnetic field acts on the particles, producing a circular path. When a particle emerges from the dee, the sign of the accelerating potential is reversed, causing the particle to speed up towards the other dee. This occurs so that a proton will accelerate towards a negatively charged dee as it exits the positively charged dee. Each time the particles cross the gap between the dees, their speed increases and they travel in a semicircle of larger radius. They gain energy with each revolution until they have sufficient energy to exit the accelerator (Figure 17.2.3).

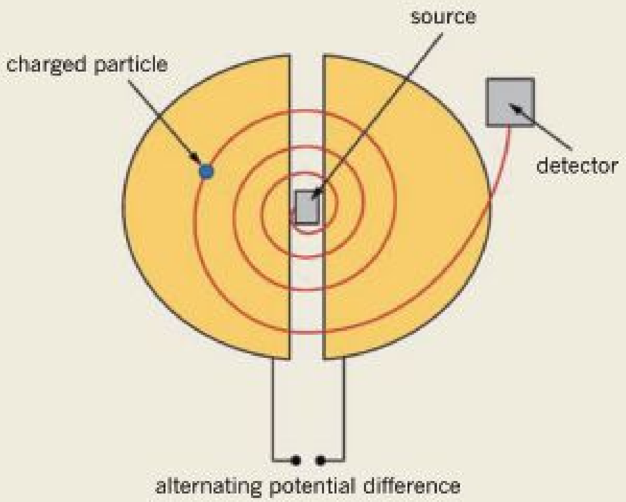


FIGURE 17.2.3 The cyclotron operates in some ways like a spiral linac—particles are accelerated from a source, through semicircular chambers called dees, until they gain sufficient energy to exit.

A key to the operation of the cyclotron is that the frequency of the radio-frequency (RF) generator that produces the alternating field must match the frequency of the circulating charged particles. The charged particles travel in a path of radius:

$$r = \frac{mv}{qB}$$

Their speed is then $v = \frac{rqB}{m}$ and the time taken for one orbit of the cyclotron is $t = \frac{d}{v}$, where d = path distance of one revolution:

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

$$= \frac{2\pi mv}{qBv} \text{ (substituting the above expression for path radius, } r\text{)}$$

$$= \frac{2\pi m}{qB}$$

Strangely enough, the time taken for one revolution of the cyclotron does not depend upon the velocity of circulating charges. This is because as the speed increases, the radius of path travelled also increases and the time taken for each orbit remains the same.

In 1943, the Adelaide-born physicist Marcus Oliphant, while working in Britain, suggested modifying the cyclotron design to produce a synchrotron.

CERN

The term CERN is derived from the French 'Conseil Européen pour la Recherche Nucléaire', or European Council for Nuclear Research. CERN was founded in 1952 to establish a world-class facility for particle-physics research in Europe. Initially concentrating on the inside of the atom, hence the word 'nuclear', CERN's main area of research is the study of the fundamental particles that make up matter and the forces acting between them. It is referred to as the European laboratory for particle physics because of this research focus. CERN started up its first synchrotron in 1957. The development of the CERN facility since then is indicative of the development of collider technology in general. Figure 17.2.4 shows how the technology at CERN has developed over time.

The first accelerator commissioned at CERN was a 600 MeV synchrocyclotron, a cyclotron in which higher energies were reached compared with a standard cyclotron. This was achieved by synchronising the accelerating voltage with the particle velocity. While in terms of energy this wasn't a significant advance over Lawrence's original cyclotron, the design of this and other synchrocyclotrons did allow for potentially much higher energy levels. The 1950s Cold War–era became something of a technological race to develop the highest energy accelerator.

After the launch of the Russian 'Synchrophasotron', a 10 GeV accelerator, CERN developed the Proton Synchrotron, first accelerating protons in November 1959. The Proton Synchrotron (PS) achieved a beam energy of 28 GeV on its first day of operation. It was the world's highest energy particle accelerator at the time. Since then, the PS has had its proton beam intensity increased to 1000 times that of the original intensity and it has accelerated many different kinds of particles besides protons. It remains in active use today, feeding accelerated particles to experiments and to other accelerators.

Particle accelerators have continued to develop in design, in the range of particles being accelerated and in the energy levels attained. In 1976, CERN switched on the 'Super Proton Synchrotron', with a beam of protons circulating a distance of 7 kilometres. It achieved 400 GeV for the first time and is now running at up to 450 GeV. In 2008, the 27-kilometre-diameter Large Hadron Collider started up. It is the largest particle accelerator ever developed, and its initial intensities of up to 8 TeV (8×10^{12} eV) were achieved, with the potential for 14 TeV being possible. This represents an increase in energy levels of almost 30 000 times greater than the levels achieved by the cyclotrons of the early 1950s.

Not all particle accelerators accelerate particles. The antiproton decelerator at CERN has been designed to slow down antiprotons, reducing their energy, to allow the study of antimatter.

Table 17.2.1 compares the advances in the design, energy level and range of particles that are accelerated as particle accelerator designs have advanced.

Commissioned	Particle accelerator	Energy level	Particles accelerated
1934	first Lawrence Cyclotron	80 eV	protons
1945	Lawrence Synchrocyclotron	730 eV	protons
1957	first CERN Synchrocyclotron	600 MeV	protons
1957	Russian Dubna Synchrophasotron	10 GeV	protons
1959	CERN Proton Synchrotron	28 GeV	protons, range of other particles
1976	CERN Super Proton Synchrotron	400–450 GeV	protons, range of other particles
1989	Large Electron Positron collider	100–209 GeV	electrons, positrons, range of other particles
2008	Large Hadron Collider	14 TeV	protons, lead ions, hydrogen ions, positrons, antiprotons

TABLE 17.2.1 A comparison of particle accelerators.

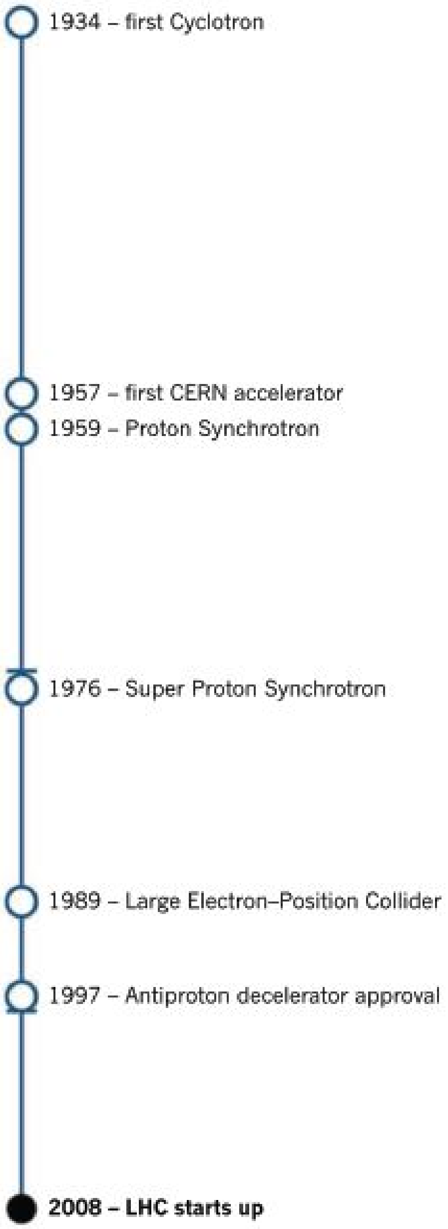


FIGURE 17.2.4 Timeline of the introduction of accelerators at CERN. The first unit deployed at CERN was in 1957, 23 years after Ernest Lawrence built the first cyclotron.

PARTICLE PHYSICS AND COLLIDERS

By the mid 1930s it was understood that atoms were made up of smaller particles and that those particles were not limited to the proton, neutron and electron. The positron, neutrino and photon, or gamma (γ) particle, had also been discovered. In 1935, Japanese physicist Hideki Yulawa predicted the existence of an additional particle in order to balance the strong nuclear force. (The strong nuclear force holds particles together in the atomic nucleus and has since been found to hold together elementary particles.) Yulawa predicted that this particle would have a mass somewhere between that of an electron and a proton, so it was called a meson, meaning 'in the middle'.

The meson that Yukawa predicted was finally discovered in cosmic rays in 1947. The search for it also led to the discovery of other particles in the interim and the development of elementary particle physics and the scientific instruments with which the interactions of particles making up an atom could be observed.

The Standard Model of particle physics, explained in Chapter 5, describes the workings of the atom as much more complicated than the simple models proposed at the beginning of the twentieth century. Research following the discovery of the meson led to not just a few more subnuclear particles, as originally expected, but many hundreds more. Colliders, and the ever-increasing particle energies that successive improvements to particle accelerators allowed, have been fundamental to these discoveries.

Recall that the hadrons are a group of particles that includes protons and neutrons. Leptons are the group of particles that includes electrons. Hadrons interact via the strong nuclear force; leptons don't and are far more numerous. While leptons are considered elementary particles, collider experiments that indicated that hadrons have an internal structure. It was discovered that hadrons are made up of more fundamental elementary point particles termed quarks. Scientists proposed that there are six quarks:

- up
- down
- charmed
- strange
- top (also called truth)
- bottom (also called beauty).

So far there is no direct evidence of up and down quarks; they may not even exist singly. Strong evidence of the truth or top quark wasn't observed until 1995. Its comparatively high rest mass required a proton–antiproton collision of almost 2 TeV in order to detect the decay products from this very-short-lived (10^{-23} s) particle. The LHCb (Large Hadron Collider beauty) detector has been specifically built to detect the bottom or beauty quark using the upgraded LHC.

All particles also have antiparticles, subatomic particles that have the same mass as the particle but with the opposite electric or magnetic properties. Many subnuclear particles are unstable and decay. Unstable particles influenced by the strong nuclear force decay more quickly than those caused by the weak nuclear force. Recording collision events in particle accelerators is essential to observing these extremely short-lived particles via observations of the particles to which they decay.

Table 17.2.2 summarises many of the subatomic particles along with their charge, rest mass and lifespan until they decay into other particles or forms. The extremely short lifespans of many of the particles makes observing them via a single collision highly problematic. This is why there are millions of collision events in a typical run at the LHC.

Category	Name	Symbol and charge	Antiparticle symbol	Rest mass ($MeVc^{-2}$) 1 $MeVc^{-2}$ = 1.783×10^{-30} kg	Lifetime (s)
Gauge boson	photon	γ	γ (self)	0	stable
	W	W^+	W^-	80.33×10^3	3×10^{-25}
	Z	Z^0	Z^0	90.19×10^3	3×10^{-25}
Leptons	electron	e^-	e^+	0.511	stable
	neutrino (e)	ν_e	$\bar{\nu}_e$	0	stable
	muon	μ^-	μ^+	105.7	2.2×10^{-6}
	neutrino (μ)	ν_μ	$\bar{\nu}_\mu$	0	stable
	tau	τ^-	τ^+	1777	2.91×10^{-13}
	neutrino (τ)	ν_τ	$\bar{\nu}_\tau$	0	stable
Hadrons: mesons	pion	π^+	π^-	139.6	2.60×10^{-8}
	kaon (also other variations)	K^+	K^-	493.7	1.24×10^{-8}
	eta	η	η (self)	547.5	5×10^{-19}
	and others				
Hadrons: baryons	proton	p	$\bar{p}$	938.3	stable
	neutron	n	$\bar{n}$	939.6	887
	lambda	Λ^0	$\overline{\Lambda^0}$	1115.7	2.63×10^{-10}
	sigma (also other variations)	Σ^0	$\overline{\Sigma^0}$	1189.4	0.80×10^{-10}
	xi (also other variations)	Ξ^-	Ξ^+	1321.3	1.64×10^{-10}
	omega	Ω^-	Ω^+	1672.5	0.82×10^{-10}

TABLE 17.2.2 Characteristics of subatomic particles and their antiparticles.

The design of particle accelerators and the associated detectors must take into account the predicted lifespan of the particle before it decays, the charge on the particle and the relative rest mass (a larger mass means that more energy is required) in order to be sure that the particle being detected meets theoretical predictions within reasonable experimental constraints.

FINDING THE HIGGS BOSON

Einstein long hoped to find a unified theory linking the four fundamental forces in nature. Theoretical work in particle physics is attempting to establish a unified theory by first developing predictions based on what is already known and then using experimental observations that could prove, or disprove, the predictions, to eventually lead to a unified model.

The Standard Model is the best model yet for linking the strong nuclear force with the electroweak theory. The **electroweak theory** unifies the weak nuclear force and electromagnetic interactions. This theory proposes that the weak nuclear force and the electromagnetic force are essentially different forms of the same electroweak interaction.

The theory has successfully predicted the W^+, W^- and Z^0 bosons. The theory also predicts that there would be a particle that, by means of a Higgs field (named after Peter Higgs, the physicist who proposed it), would allow for the considerable masses of the W and Z bosons rather than them having no mass like a photon. The Higgs boson would interact with the W and Z bosons to force them to go slower than the speed of light and hence acquire mass, essentially giving mass to all elementary particles.

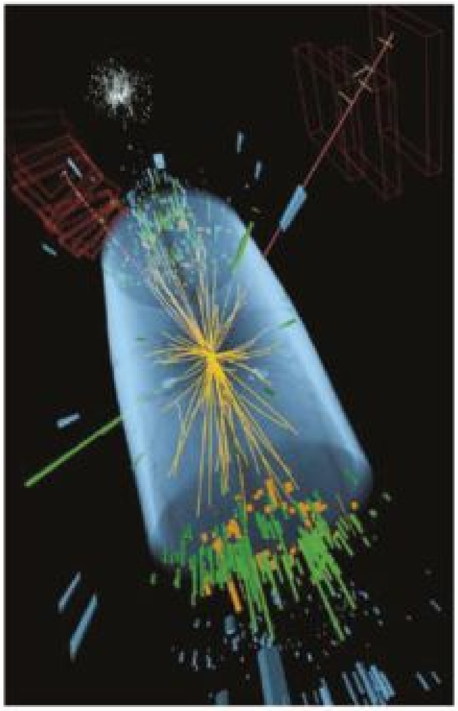


FIGURE 17.2.5 A Higgs boson event resulting from collisions between protons in the CMS detector in the LHC. The particle decays into two bosons, one of which decays to a pair of electrons and the other to a pair of muons.

PHYSICSFILE

Limitations of the Standard Model

The Standard Model is a very powerful theory and is well supported by experimental evidence. However, some phenomena, such as dark matter and the absence of antimatter in the universe, can't be accounted for in the model.

On the 4 July 2012, experimenters working with the ATLAS and CMS detectors at the LHC announced that they had observed a new particle with a mass in the region of 126 GeV, which was consistent with the predicted Higgs boson (see Figure 17.2.5). Excitingly, on the 17 March 2015, the final results were presented: a mass for the Higgs boson of 25.09 ± 0.24 GeV, corresponding to a precision of better than 0.2%. This demonstrates the importance of particle colliders in the high-precision confirmation of theories in particle physics.

THE LARGE HADRON COLLIDER

The **Large Hadron Collider** (LHC) at CERN is one of the latest developments in particle accelerator technology. Due to its enormous size and high energies, the particle accelerator is probably the best known outside of particle physics research. However, the LHC is only the most recent and highest energy accelerator in a chain of accelerators and experiments that work together to feed particle research activities at CERN. Together, the particle detectors working from the LHC allow specialised investigations at the leading edge of particle-physics research.

Facts and figures about the LHC:

- It gets its name from: its size (Large); its ability to accelerate both protons and ions, both of which are hadrons (Hadron); and because it is a collider in which two particle beams travelling in opposite directions collide at points where the two rings of the LHC intersect (Collider).
- It is the largest scientific instrument ever built: the tunnel has a circumference of 27 km with an average depth of 100 m below ground level (see Figure 17.2.6). Being located underground has the extra benefit of shielding the accelerator from external radiation, although cost was the main factor for building underground.
- The LHC was initially considered in the 1980s with the intent of producing an accelerator capable of reaching the (predicted) energies needed for research on particles that would have existed just 10^{-12} seconds after the big bang.
- It was built at a cost of €3 billion (equivalent to around AUD$4.5 billion), with financial contributions from non-member countries (i.e. Japan, India and the USA) helping to reduce the construction time.
- The project was built for an initial four experiments (ATLAS, ALICE, CMS and LHCb). Two additional experiments have since been added.
- Its first collisions in October 2008 produced particles with a (then record) energy of 4 TeV. After a two-year upgrade, it is now capable of energies potentially reaching 14 TeV or 7 TeV per proton beam in a head-on collision. Lead-ion beams now have a collision potential of 1150 TeV.
- Protons complete over 11 000 revolutions of the 27 km circumference each second, travelling at a speed of 0.999999991 times that of light. There are up to 600 million collisions per second.

FIGURE 17.2.6 A section of the LHC's 27 km tunnel.

The main steps in the acceleration of particles in the LHC are outlined below.

- Hydrogen atoms are pumped from storage and are stripped of their electrons to create protons.
- The protons travel through a linac reaching an energy of 50 MeV and are injected into the proton synchrotron booster.
- The proton synchrotron booster accelerates the protons to 1.4 GeV, after which they are fed into the proton synchrotron and accelerated to 25 GeV.
- Protons are then sent to the super proton synchrotron and are accelerated to 450 GeV.
- Finally, the protons are injected into the LHC in both clockwise and anticlockwise directions to create the two beams that will finally collide.
- The beams are accelerated to 7 TeV each over a period of 20 minutes, after which they can be kept in the storage rings for a number of hours.

Goals of the LHC

The Standard Model summarises the current understanding of particle physics. It has been tested and has been successful in predicting previously unknown particles. However, it is unable to answer questions on dark matter and why there is no antimatter in the universe, among others. The main goals of the LHC aim to add to the understanding of particle physics beyond the Standard Model. Some examples are given below.

- Bosons are also referred to as force-carrying particles and are used to explain forces being exerted on particles by fields. Scientists are seeking a unified theory for the four natural forces. The Standard Model links the weak nuclear force, the strong nuclear force and the electromagnetic force, but is unable to construct a similar theory for gravity. Supersymmetry, the existence of more massive particles than are currently known, could lead to a unified theory.
- The reasons why objects and particles have mass cannot be predicted from the Standard Model. A separate model is needed. The Higgs field is theorised to interact with particles such that they acquire their mass. The Higgs field has a Higgs boson particle associated with it. The successful detection of the Higgs boson and subsequent determination of its mass was a major achievement of the LHC.
- Astronomical observations suggest that only 4% of all of the matter in the universe is visible. The LHC is searching for evidence of dark matter and dark energy, which are theorised to account for 23% and 73% of the remaining matter respectively.
- Matter and antimatter existed together in equal quantities at the time of the big bang but today, as far as we know, there is no antimatter. Experiments at the LHC will attempt to determine why.
- Heavy ions such as lead (Pb) colliding at high energies form hot, dense matter. The LHC will be used to investigate the state of matter called 'quark–gluon plasma' that is theorised to have existed in the early universe.

The initial four experiment stations developed as part of the LHC project are described below.

CMS

The CMS or Compact Muon Solenoid, shown in Figure 17.2.7, is one of two general-purpose detectors at the LHC. It is the centre of one of the largest scientific collaborations ever, with 4300 scientists from 182 institutions in 42 countries working with CMS data. It was used in the search for the Higgs boson and is being used to search for extra dimensions and dark matter.

The CMS differs substantially from the other general-purpose detector in its magnet system. A steel yoke, making up the majority of the detector's 14 000 tonne mass, confines a magnetic field of 4 tesla (the Earth's magnetic field is around 50 microtesla) generated by a cylindrical coil. A cylindrical coil that carries current to create a magnetic field is called a solenoid, hence the CMS name.

FIGURE 17.2.7 The CMS (Compact Muon Solenoid) is a general purpose detector at the LHC (Large Hadron Collider). The worker inside the central coil highlights the size.

ATLAS

The 7000 tonne ATLAS particle detector is the other general-purpose detector, along with the CMS, analysing collisions from the LHC. ATLAS, short for 'A Toroidal LHC ApparatuS' detects head-on, proton–proton collisions. While lighter than the CMS, in size it is the largest particle detector ever built.

Being a general detector, it is able to investigate a wide range of physics theories. It was used alongside the CMS detector in the search that found the Higgs boson, and is being used for experiments in detecting extra dimensions and the search for particles that could potentially lead to the proof of the existence of dark matter.

The scientific goals that led to the development of ATLAS are the same as those for the CMS experiment, but ATLAS utilises different technical solutions and has a different magnet-system design. In May 2015 it detected the first proton–proton collisions based on the increased 13 TeV energies of the upgraded LHC (see Figure 17.2.8).

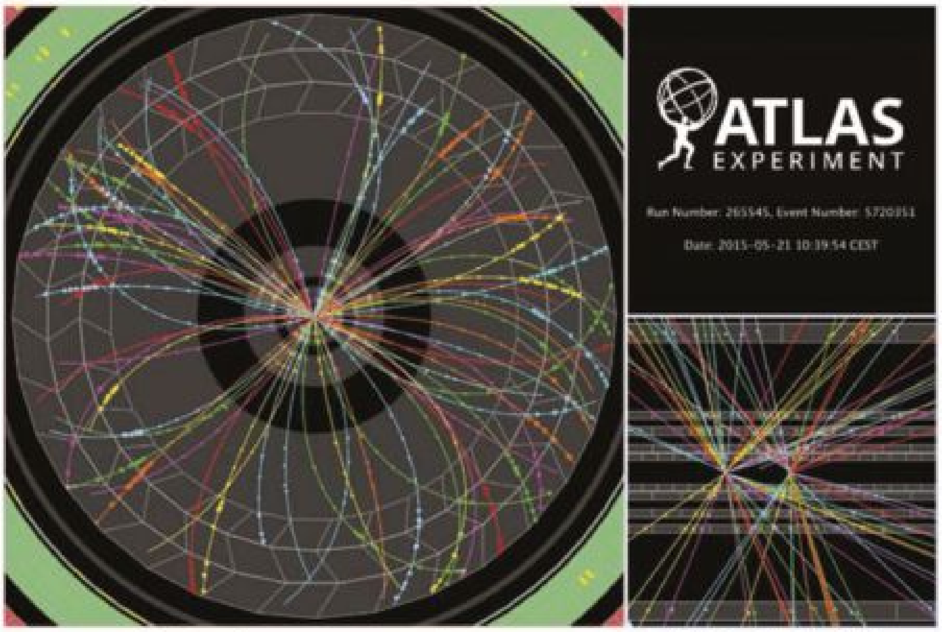


FIGURE 17.2.8 Displays of a proton–proton collision event recorded by ATLAS on 21 May 2015, at collision energy.

A key design feature that enables the ATLAS to act as a general-purpose detector and differentiates it from the CMS experiment is the toroidal-shaped magnet system at its centre. The magnet consists of eight 25 m long superconducting coils forming a cylinder around the beam pipe.

ALICE

Particle physicists theorise that hadrons, including protons and neutrons, are made up of quarks, bound together by particles called gluons. While physicists are fairly certain quarks exist (previous research supports the theory for their existence), no quark has been identified in isolation. They seem only to exist permanently bound together in hadrons.

When matter is subjected to extreme temperature conditions simulating those just after the big bang, matter changes phase to a form called quark–gluon plasma. ALICE, 'A Large Ion Collider Experiment', is a heavy ion detector designed specifically to study matter as it interacts under extreme temperatures (see Figure 17.2.9). Collisions of heavy ions in the LHC generate temperatures of 5.5×10^{12} kelvin. At these temperatures protons and neutrons break down, freeing quarks from the gluon bonds to form the quark–gluon plasma. The 10 000 tonne ALICE detector studies the plasma as it initially expands and then cools, looking for the quarks that make up matter.

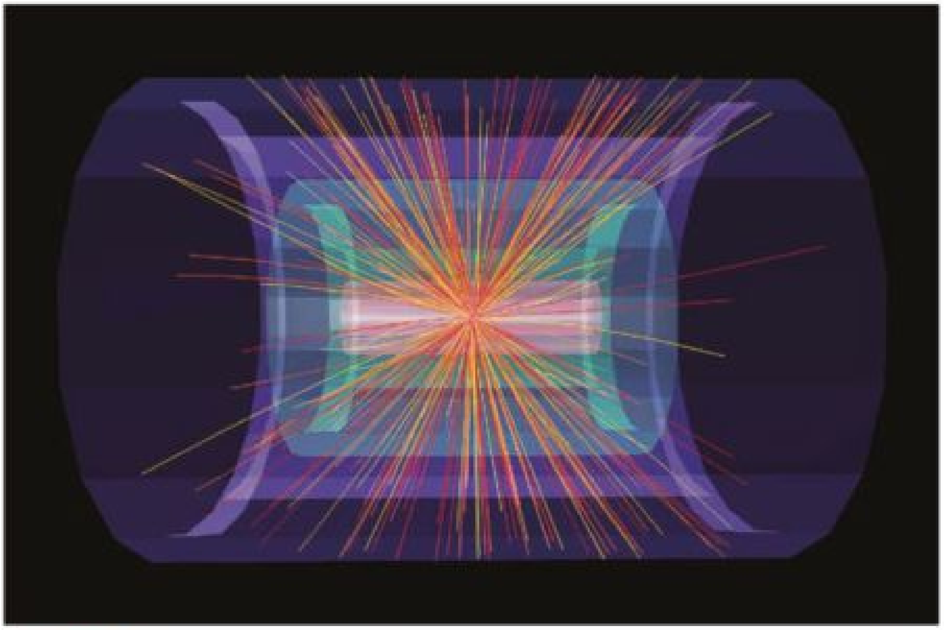


FIGURE 17.2.9 Particle tracks from the first lead-ion collisions seen by the ALICE detector at CERN. Lead-ion collisions are expected to produce quark–gluon plasma, a primordial state of matter thought to have been present in the universe microseconds after the big bang.

ALICE utilises two superconducting magnet systems, bending charged particles so that their momentum can be measured. An inner solenoid generates a 2 tesla magnetic field. An outer toroidal-shaped magnetic field is produced by 8 superconducting barrel loops and two end-cap toroidal magnets. ALICE generates 1 petabyte (10^{15} bytes) of raw data per second during collision detection.

LHCb

LHCb stands for 'Large Hadron Collider beauty', an experiment designed to investigate the slight differences between matter and antimatter. It is shown in Figure 17.2.10. It was specifically built to detect the bottom (or beauty) quark, also called the b quark, which is a frequent decay product for the Higgs boson. By studying the b quark, scientists hope to better understand what happened immediately after the big bang.

According to the theory, approximately 14 billion years ago the big bang produced both matter and antimatter. As the matter cooled and expanded its composition changed. Within one second, it is believed that antimatter had largely ceased to exist, leaving matter to create the universe as you know it. The b quark, an antiparticle, is believed to be key in understanding that process.

Every collision event at the LHC forms a broad range of quarks. The ATLAS and CMS detectors enclose the entire collision point within a detector. By contrast, the LHCb is able to detect b quarks by having a series of sub-detectors at points largely forwards of the collision point in order to focus on particles thrown forwards of the collision. The first sub-detector is close to and forwards of the collision event; the rest of the chain of sub-detectors are arranged one behind the next, over a total length of 20 metres.

Since the quarks created by the collision event decay quickly into other subatomic particles, the LHCb features movable tracking detectors consisting of a forward spectrometer and planar detectors.

Weighing 5600 tonnes, the LHCb is sited 100 metres below ground. In May 2015, when the LHC was reopened after the upgrade to energies of 13 TeV, the LHCb detected events from proton–proton collisions. These initial collisions were only in the 450 GeV range and hence not suitable for full physics studies, but were able to confirm that the detector was recording suitable events. Follow-up studies in June 2015 were among the first to use the increased energy of the LHC.

FIGURE 17.2.10 The underground cavern holding the LHC beauty, or LHCb, experiment of the Large Hadron Collider at CERN. Much of the work on this experiment is directed to the study of antiparticles.

17.2 Review

SUMMARY

- Colliders are used to collide two beams of particles head on. They were developed for the study of elementary particle physics.
- Ernest Lawrence, was the first to produce a working cyclotron in 1932. This first cyclotron was just 13 cm in diameter and accelerated protons to 80 keV (1 keV = 1.6×10^{-16} J).
- As the study of particle physics rapidly progressed in the 1950s, so did the energies developed by successive generations of particle accelerators. The LHC can now achieve energy levels of 14 TeV, almost 30 000 times that achieved in the early 1950s.
- Particle accelerators have also developed to be able to accelerate a range of particles, including protons, electrons, antiprotons, positrons and ions, to allow specific experiments in particle physics.
- Not all particle accelerators accelerate particles. The antiproton decelerator at CERN has been designed to slow down antiprotons to allow the study of antimatter.
- The design of particle accelerators and the associated detectors must take into account the charge on the particle being investigated, the relative rest mass (a larger mass means that more energy is required) and the predicted lifetime of the particle before it decays in order to be sure that the particle being detected meets theoretical predictions within reasonable experimental constraints.
- Observations of the Higgs boson at the LHC confirmed a prediction of the electroweak theory and provided observational evidence for the Standard Model.
- The LHC gets its name from its size, its ability to accelerate both protons and ions, both of which are hadrons, and because it makes two particle beams travelling in opposite directions collide at points where the two rings of the LHC intersect.
- The main goals of the LHC aim to add to the current understanding of particle physics beyond the Standard Model.
- The LHC's four initial experiments differ in design and/or purpose:
  - CMS, or Compact Muon Solenoid, is one of two general-purpose detectors. Its magnets form a large solenoid producing a magnetic field of 4 tesla. It was involved in the discovery of the Higgs boson.
  - ATLAS, short for 'A Toroidal LHC ApparatuS' detects head on proton–proton collisions. It is the other general-purpose detector and was also involved in the search that found the Higgs boson. It has a toroidal, or doughnut-shaped, magnet system at its centre.
  - ALICE, A Large Ion Collider Experiment, is a heavy-ion detector designed specifically to study matter as it interacts under extreme temperatures. Two superconducting magnet systems bend charged particles so that their momenta can be measured.
  - LHCb stands for Large Hadron Collider beauty (LHCb), an experiment designed to investigating the slight differences between matter and antimatter. It does so via a series of sub-detectors at points largely forwards of the collision point in order to focus on particles thrown forwards of the collision.

KEY QUESTIONS

1 How did cyclotrons get their name?

2 How is the synchrocyclotron an advance on the earlier cyclotron designs?

3 The Higgs boson was theorised to interact with other bosons, W and Z in particular, forcing them to go slower than the speed of light. Describe the essential role of the Higgs boson.

4 The Standard Model is a very powerful theory and is well supported by experimental evidence from particle accelerators. It can't explain some predicted phenomena that are the subject of current collider experiments. Which of the following are not explained by the standard model?
Higgs boson, dark matter, strong nuclear force, antimatter, weak nuclear force

5 The Large Hadron Collider, or LHC, accelerates particles from a proton or hadron source. It is the largest scientific instrument ever built. Place the following components of the LHC in order from source to detector:
storage ring
proton synchrotron booster
electrons stripped away
detectors
hydrogen atom source
super proton synchrotron
linac

6 The LHCb detector has been designed to investigate differences between matter and antimatter. What design feature is incorporated into the LHCb to enable this form of investigation?

7 A proton–proton collision in the Large Hadron Collider can generate 14 TeV energies. Using 1 eV = -1.6×10^{-19} J, calculate the energy produced by the accelerator in joules.

8 Which particle collision is the focus of the ALICE detector at the LHC?

9 Briefly explain the role of the proton synchrotron booster in the LHC complex. Include estimates of the relevant particle energies in your answer.

10 LHCb stands for Large Hadron Collider beauty (LHCb). Explain its role as a detector at the LHC and the design feature that enables it to fulfil its role.

17.3 The importance of accelerator technology in society

Particle accelerators represent an immense investment in time and resources. The Australian Synchrotron initially cost $221 million to build and has a running cost of around $25 million per year. The Large Hadron Collider is one of the most expensive scientific instruments ever constructed. It cost an estimated $8.5 billion to build and its operating budget is in excess of $1.3 billion per year. Even with these costs, countries continue to willingly invest in these technologies.

The economic benefits from the discoveries made by these instruments are many times greater than the costs. The Australian Synchrotron is expected to generate economic benefits approaching $21.7 billion over its 25-year operating life. While the discovery of the Higgs boson may seem to have little to do with the lives of average people, the regular day-to-day discoveries are having a significant impact on daily lives and will continue to do so into the future.

From the development of new pharmaceuticals that treat major diseases to the ground work behind the development of new antibiotics and the treatment of pollution, more than 30 000 high-energy particle accelerators around the world provide the tools for scientists to understand structures and particles that would otherwise be beyond their means to investigate.

LARGE HADRON COLLIDER—DATA ANALYSIS

The previous section described the work of the LHC at CERN. Particles are colliding within the LHC approximately 600 million times per second in ATLAS and CMS detectors, interacting in complex ways and generating enormous amounts of data. The sheer volume of data is extraordinary:

- Each collision event generates, in round figures, 1 MB of data. That means, operating normally, the amount of data generated per second equals 600 million MB or 600 TB per second. If the average home computer has a 1 TB hard drive, one second of data is sufficient to fill the capacity of 600 hard drives.
- Allowing for two ten-hour shifts per day over 300 working days per year, the total volume of data per year is approximately 600×10^{12} bytes per second × 3600 (bytes each hour) × 20 (hours per day) × 300 (days per year) equates to 1.3×10^{22} bytes or 13 zettabytes of raw data each year.

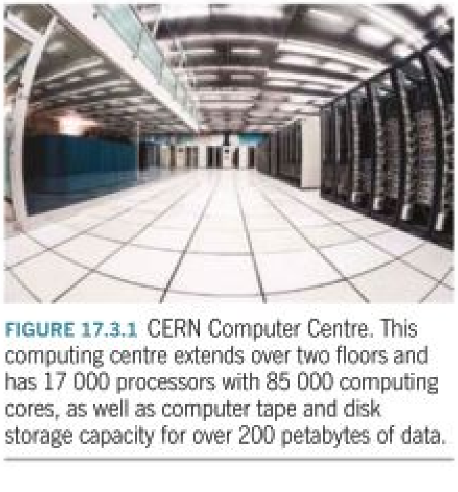


FIGURE 17.3.1 CERN Computer Centre. This computing centre extends over two floors and has 17 000 processors with 85 000 computing cores, as well as computer tape and disk storage capacity for over 200 petabytes of data.

This information must be initially stored in the computing centres like the one in Figure 17.3.1, then analysed so that events of interest can be extracted for comparison to theoretical models.

The CERN computer centre's 17 000 processors, 85 000 computing cores and over 200 petabytes of data storage allow initial storage and analysis of the data. CERN retains around 25 petabytes of that data per year. The remainder of the data is shared among computing centres around the world via the 'Worldwide LHC Computing Grid'.

The grid is divided into 4 tiers. Tier 0 is the CERN data centre itself. Once experimenters at CERN determine which collisions have yielded data that is of genuine interest in furthering their studies, one complete copy of the initial raw data is stored there. The same data is then shared between 11 Tier 1 data centres across Europe, Asia and North America.

Tier 1 centres then distribute the data to any of around 140 Tier 2 centres that express interest in particular collision events. Tier 2 centres are largely universities, where researchers can use the data to add to or contrast with experiments of their own. The data for these events can be stored permanently at the university as required.

Tier 3 applies to individual scientists or researchers. Any scientist who collaborates on a project run from CERN is able to access the data by placing a request with CERN or a Tier 1 centre. These scientists are also able to distribute data to their own students for further analysis.

Since November 2014, a further level of distribution has been added. Around three years after the initial collision event and after the data has been accessible through the Worldwide LHC Computing Grid, the collision data is made available through the CERN Open Data Portal. Data sets from the ALICE, ATLAS, CMS and LHCb experimental structures are converted into formats that can be readily visualised and analysed by open-source software specifically developed for the purpose. The full set of data is then made available through the specially developed CERN website under open-source licences. Unique identifiers allow the data sets to be cited and referenced. This unique approach to the distribution of the huge volumes of data generated by the LHC means that anyone is able to analyse data that could potentially lead to historic discoveries.

The CERN Open Data Portal provides enormous opportunities for advancements in areas as diverse as environmental science, particle physics and even treatment for concussion. In past years, the enormous amount of data generated by the LHC was available only to organisations as part of the Worldwide LHC Computing Grid. But now, data is available to all scientists involved in fields of related research. Many years can be saved on fundamental research, leading to drastically reduced times between theory and development.

APPLICATION OF PARTICLE ACCELERATORS

Particle accelerators aren't just used as a research tool for particle physicists. The Large Hadron Collider got so much media attention with the proof of the existence of the Higgs boson in 2012 that you'd think that was its sole purpose. These large particle colliders will continue to be developed into the future as scientists try to produce experimental evidence for new theories and ideas. Large-scale developments of the LHC will see it continue to play a leading role well into the future.

Particle accelerators take many forms—from synchrotrons like the Australian Synchrotron at Monash University, to relatively small-scale particle accelerators used for implanting ions in silicon wafers (Figure 17.3.2), and even the humble cathode ray tube found in old-style televisions. This range of particle accelerators can be used for a wide range of applications and for the development of diverse technologies.

FIGURE 17.3.2 The interior view of an opened ion accelerator.

Other areas where particle accelerator research is leading to ground-breaking discoveries include (but are not limited to):

- cancer treatment
- carbon dating
- art restoration
- the production of artificial body parts.

Believe it or not, X-ray microscopy has even led to better nappies!

Pharmaceutical research

The rise of synchrotron light sources has enabled scientists to discover previously unknown details about the structure of molecules. In 1960, fewer than a dozen protein structures were identified down to the detail of individual atoms. By May 2015, nearly 109 000 structures have been deposited in the Protein Data Bank, with more being revealed all the time.

Armed with a better knowledge of molecular structures, scientists can design drugs to inhibit the function of agents that cause disease. For example, the influenza virus relies on an enzyme called neuraminidase to spread it from cell to cell. If this could be inhibited then the virus would not be able to spread through its host.

In 1978, Dr Graeme Laver of the Australian National University's John Curtin School of Medical Research successfully crystallised neuraminidase. Dr Peter Colman and Dr Jose Varghese from CSIRO then spent years trying to unlock the structure of the protein. They used synchrotron sources from Tsukuba, Hamburg and Stanford to carry out X-ray diffraction experiments. Synchrotron data is particularly well suited to X-ray diffraction studies as it is high energy, highly collimated and can be produced continuously for several hours. The CSIRO team solved the structure of neuraminidase, shown in Figure 17.3.3, in 1983. Once the structure was revealed, they identified a particular section of the enzyme that appeared the same across all strains of influenza.

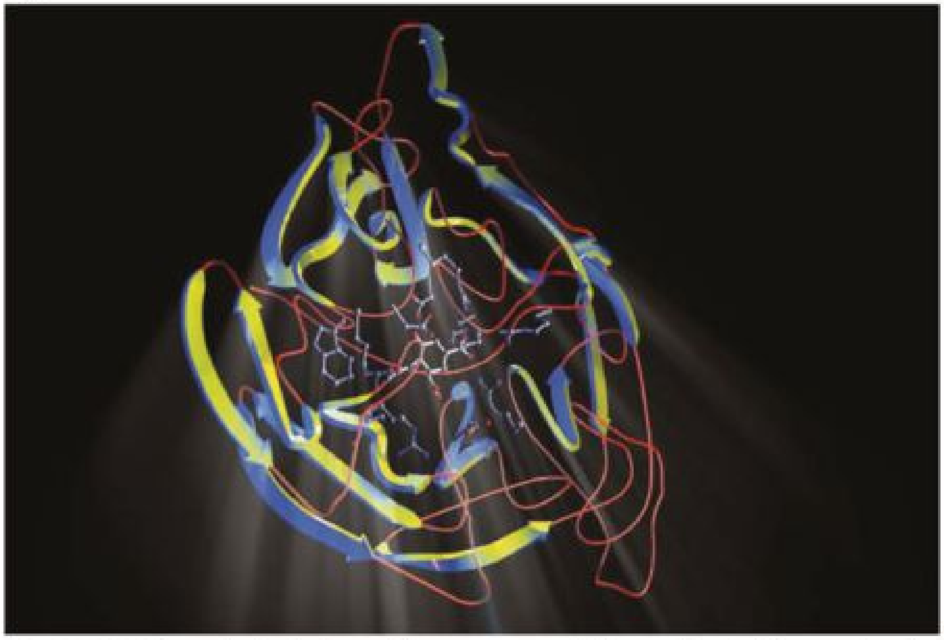


FIGURE 17.3.3 The molecular structure of the enzyme neuraminidase, which the influenza virus relies on to spread from cell to cell. The structure was determined using X-ray diffraction experiments.

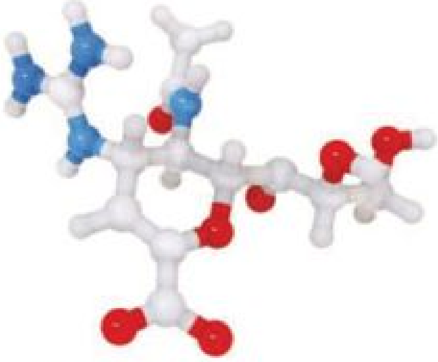


FIGURE 17.3.4 The molecular structure of the Relenza molecule (zanamivir).

Professor Mark von Itzstein, then at the Victorian College of Pharmacy at Monash University, worked with scientists from the firm Biota Holdings to design a drug that could lock into this particular section of neuraminidase and stop the virus from replicating. The drug they produced is called 'Relenza' (zanamivir), the world's first anti-influenza drug (pictured in Figure 17.3.4). It has now been approved for therapeutic use in over 64 countries.

Relenza is particularly used to treat members of the community in high-risk groups, such as the elderly and those with chronic respiratory problems, diabetes or cardiovascular disease.

Materials modification: Lithography

Miniaturisation is leading to exciting developments in many areas but none more exciting than in the fields of medicine, micro-sensors and micromachining. **Lithography** is one process with which micro devices can be produced.

Lithography is the process of printing a pattern onto another surface. While forms of this technique have existed for centuries, the method of using high-energy X-rays to etch a pattern is much more recent. Lithographic techniques can be used to create tiny mechanical parts or to produce moulds to be filled with a range of other materials, such as glass and ceramics, to make components. Since X-rays from synchrotron light are high intensity, can be produced for several hours and can be precisely directed, synchrotron light is ideally suited to this application. Using synchrotron light, lithographic techniques are capable of producing high-resolution images between 100 μm and 2 mm thick.

Micro-electrical mechanical devices are one particular example of this process. X-rays of wavelength 0.01 to 1.0 μm are shone through a patterned mask onto a **photo resist** coating that is later developed to reveal the image. This process is shown in Figure 17.3.5.

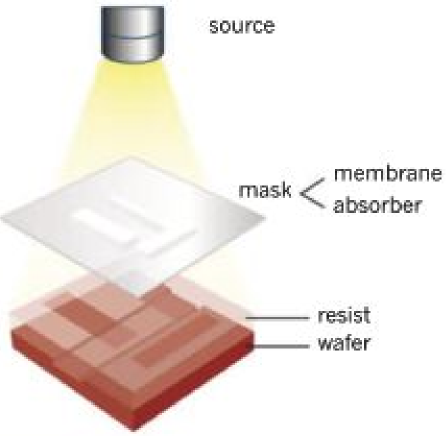


FIGURE 17.3.5 The basic set up of an X-ray lithographic process.

As the radiation is exposed, it casts a shadow of the mask pattern onto the resist. X-rays interact with the resist coating. This photosensitive layer produces images with a resolution of just 0.14 μm. The exposed resist can be removed by chemical development, leaving behind the unexposed areas.

The collimation of the X-ray beam from synchrotron light reduces the effect of blurring of the unexposed areas. Because the beam is of high intensity, shorter exposure times can be used which prevents overheating of the mask arrangement. The availability of synchrotron light has resulted in the fabrication of much smaller, more-compact storage rings in synchrotrons specifically for the performance of X-ray lithography work.

A process of deep-etch lithography, called 'LIGA', can be used to produce gears only a few thousandths of a millimetre in size (see Figure 17.3.6). These components are used in the building of micro-electronic devices in the emerging field of micromachining.

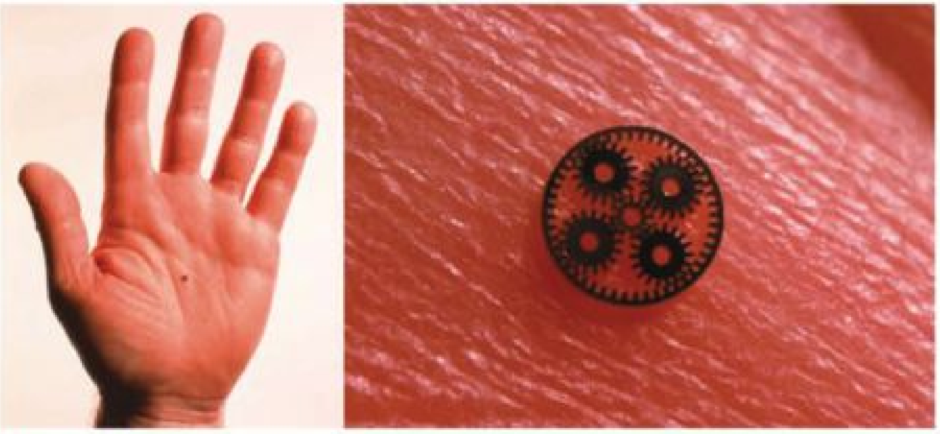


FIGURE 17.3.6 A hand holding microcogs that make up a microgear mechanism.

Semiconductor ion implantation

Semiconductors are the basis of most electronics. From mobile phones to heart pacemakers, the production of silicon wafers, allowing the miniaturisation of electronic circuits, has transformed the modern world.

A **semiconductor** is a material that will conduct electricity only under particular conditions, placing its properties somewhere between those of a conductor and an insulator. Controlling when a semiconductor will act as a conductor or as an insulator allows semiconducting materials to be used in circuits as a control or switching mechanism.

Typically, semiconductors consist largely of silicon or germanium which is 'doped' with impurities such as antimony, arsenic and metal oxides to create a device where ions carry the current. N-type semiconductors carry current in the form of negatively charged ions, and P-type semiconductors carry current as a 'hole' or electron deficiency creating a positively charged ion. The conductance of the semiconductor will then depend upon the current or voltage applied, or the intensity of any incident electromagnetic radiation. The resulting 'chip' forms the control centre of an electronic device.

Particle accelerators are widely used as a particularly effective means of implanting ions within a silicon or germanium wafer (see Figure 17.3.7). It is, in fact, the largest use of industrial particle accelerators. During the production process the semiconductor wafer is exposed to the beams from particle accelerators. The highly focused beam of a particle accelerator implants ions in very specific locations and concentrations, creating areas of high electron concentration (N-type) and areas of electron depletion or holes (P-type). The result is the production of semiconductors with a much higher density within a given size enabling manufacturers to further miniaturise devices or build more power into a similarly-sized chip.

Since ion-implantation accelerators were first used in the 1970s, there have been more than 12 000 ion-implantation accelerators installed and in operation in electronics factories around the world. An estimated 300 new units are being added each year. However, these particle accelerators have more in common with a cathode ray tube than with the Large Hadron Collider.

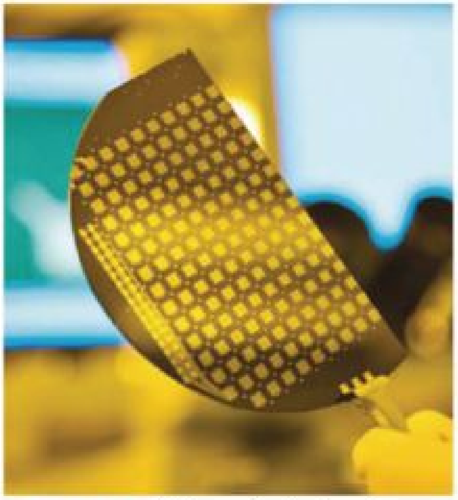


FIGURE 17.3.7 A silicon wafer that has been implanted with ions.

Small and specific to the task, the design of these particle accelerators is increasing in precision as semiconductor manufacturers work on new processes that require precise and uniform doping. A particle accelerator is particularly suited to these tasks as it provides a stable, consistent, highly collimated beam of ions. Designs allow ion currents from the 1 μA range through to 100 mA and ion energies from 100 eV through to around 10 MeV.

Due to the range of energies and ion doses required for specific applications, ion-implantation accelerators aren't just one particular design. Four broad classifications for different machine types exist:

- medium current: ion beam currents between 10 μA and 2 mA
- high current: ion beam currents between 30 mA and 100 mA
- high energy: ion energies between 200 keV and 100 MeV
- very high dose: implantation of doses greater than 10^{14} ions mm^{-2}.

For each machine classification the actual beam requirements also differ. While early machines were only designed to allow for implantation on semiconducting wafers 75 mm in size, machines introduced from 2015 allow for wafers of up to 450 mm in diameter with a dose accuracy of 1% or better and the ability to control the incident angle to within 0.5°.

Implantation particle accelerators can also be grouped according to design. Medium current beamline systems differ in basic design but all contain the same basic components (shown in Figure 17.3.8):

- a heated **cathode** acting as an ion source
- a magnetic field used for selecting the particular ions required
- an accelerator column for increasing the energy of the ions to required levels
- a collection and delivery mechanism ensuring that the delivered dose of ions can be measured, placed and stopped as required.

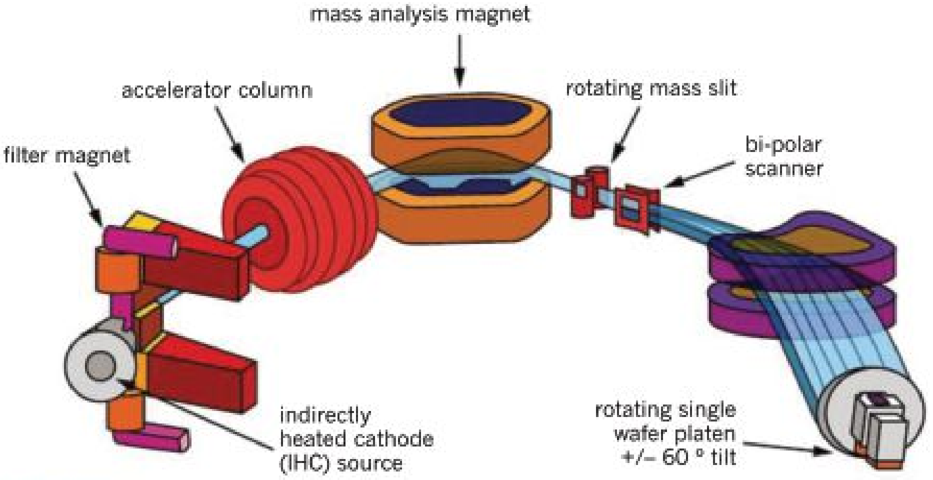


FIGURE 17.3.8 The main components of a beamline-based ion-implantation accelerator, in this case a medium-current ion implanter.

The final design of a beamline system takes into account the requirements of the particular process used. Additional components may be added to bring the beam parallel and close to the surface of the wafer target.

Other types of ion-implantation particle accelerators include high-current beamlines designed to operate at low energies and with fewer beamline components than medium current systems. These are called accel–decel beamlines (or accelerate–decelerate beamlines) since ions are only accelerated at the source extraction stage. A deceleration stage to reduce the ion energies is added after the mass-resolving slits producing ion beam energy levels as low as 0.2 keV. Other types, called MeV beamlines, are specifically developed to produce very high-energy ions in the order of 200 keV into the MeV range. Further types will continue to be developed as new applications for semiconductors are explored and brought into production.

17.3 Review

SUMMARY

- The Large Hadron Collider at CERN generates 1 MB of data per collision event which translates to 1.3×10^{22} bytes of data per year.
- The CERN computing centre processes the initial data and retains 25 petabytes of data per year. The remaining data is distributed to computing centres through the Worldwide LHC Computing Grid.
- Additional access to LHC data is available to scientists, researchers and students through the CERN Open Data Portal.
- Particle accelerators are being used for an increasing number of applications both now and into the future including (but not limited to):
  - pharmaceutical research
  - materials modification
  - semiconductor ion implantation
  - cancer treatment
  - carbon dating
  - art restoration
  - production of artificial body parts
  - environmental monitoring.

KEY QUESTIONS

1 Approximately how much data does one collision event at the LHC (Large Hadron Collider) generate?

2 What are the CERN computing centre processors used for?

3 Synchrotron light is particularly well suited to X-ray diffraction investigations of protein structures. Why is this?

4 Briefly describe what is meant by 'lithography'.

5 What attributes of synchrotron light make it suitable for lithographic processes?

6 Beamline particle accelerators are a type of medium-current accelerator used in semiconductor production. What is the source of ions in a beamline accelerator?

7 A high-energy ion-implantation particle accelerator can produce up to 100 MeV. Calculate the energy produced by the accelerator in joules.

8 What is Relenza?

9 The development of Relenza was based on understanding the structure of proteins, using which process?

10 What is a semiconductor?

Chapter review

17

KEY TERMS

anode
beamline
booster ring
cathode
cathode ray tube
cyclotron
electroweak theory
Higgs boson
interference
Large Hadron Collider
linac
lithography
particle accelerator
photo resist
semiconductor
storage ring
synchrotron
synchrotron light
X-ray diffraction

1 Where in a cathode ray tube are electrons released from?

2 What is linac the shortened term for?

3 What is a beamline?

4 Synchrotron radiation is described as having a high degree of collimation and very low beam divergence. Which particular experiment technique is this characteristic of synchrotron radiation ideal for?

5 In which section of the Australian Synchrotron do electrons move from a heated filament?

6 High-energy electrons travelling in a curved path under the influence of a magnetic field will emit light. Within what frequency range is this light? (Give a qualitative answer only.)

7 A high-energy ion-implantation particle accelerator can produce up to 200 keV. Calculate the energy produced by the accelerator in joules.

8 Successive generations of particle colliders have produced particles of ever higher energies. High energy collisions enabled scientists to determine that some subatomic particles have an internal structure. Which of the following particles have an internal structure: leptons, hadrons, quarks, bosons? Give a reason for your answer.

9 Strong experimental evidence for the truth or top quark wasn't observed until 1995. What physical constraint of particle accelerator design would have been the main limiting factor that meant the top quark couldn't be observed earlier?

10 Each of the four main experiments that form the detectors of the LHC has a particular experimental goal that influences its design. List each experiment and a description of the experimental goal of each one.

11 The LHC is 27 km in circumference. At maximum particle acceleration, protons complete 11 000 circuits per second. Using your understanding of motion, calculate the speed of the particles.

12 In Run 2 of the LHC, particles are accelerated to energies of up to 7 TeV. However, collision events are quoted as being up to 14 TeV. Explain, in terms of the collider design and features, why the quoted energy is double that of the individual particles.

13 The Standard Model of particle physics seeks to unify theories about the weak nuclear force, electromagnetic force and the strong nuclear force. It has been particularly successful in predicting particular particles. Name two areas where the Standard Model is deficient.

14 In 1960, fewer than a dozen protein structures were identified down to the detail of individual atoms. Approximately how many structures have been registered with the Protein Data Bank as of May 2015?

15 Describe the ion-implantation process and why it is advantageous.

16 Which type of machine is heavily involved in ion implantation?

17 Write down the core components of a beamline ion-implantation accelerator in the correct order from ion source to implantation.

18 Why does CERN need such a powerful computer centre?

19 The Worldwide LHC Computing Grid comprises three tiers of data processing centres and general research and science use. Where do universities fit into the tier system?

20 What is the Open Data Portal within the tier system at CERN and who can access it?

CHAPTER 18 Sport

Sport is an important part of Australian culture. Many Australians seem to have a particular love for ball sports such as cricket and football. In recent years, competitors have turned to science to help them improve their performance in all areas of their chosen sports.

An understanding of physics concepts such as friction, momentum and energy can help to predict the motion of a ball as it moves through the air and bounces off the ground. It can also allow the motion of the ball to be controlled.

Key knowledge

By the end of this chapter you will have studied the physics of sport, and will be able to:

- investigate and calculate theoretically and practically the transfer of momentum in elastic and inelastic collisions (limited to two dimensions) including the use of the coefficient of restitution, e
- investigate and apply theoretically and practically the coefficients of static and kinetic friction to sliding and rolling balls to calculate speeds using Newton's laws of motion and the equations of constant acceleration
- explain rolling of spherical objects using angular and linear speeds: $v = r\omega$
- model and describe qualitatively the energy transfers in the action of a double pendulum in at least one of the following:
  - the swing of a racquet, club, stick or bat
  - the throw, pitch or hurl of a ball
  - the kick of a ball
- calculate air resistance (drag) and terminal velocity: $F_D = \frac{1}{2}C_D\rho v^2 A$
- investigate and apply theoretically and practically the equations of constant acceleration to calculate the flight of objects through the air (neglecting air resistance) in two dimensions
- model and describe qualitatively the flight of:
  - a ball through the air when air resistance is not neglected
  - spinning sports balls with reference to the Magnus effect
- analyse and explain the relative influence of dynamics factors that affect the performance of equipment in ball sports.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

18.1 Collisions

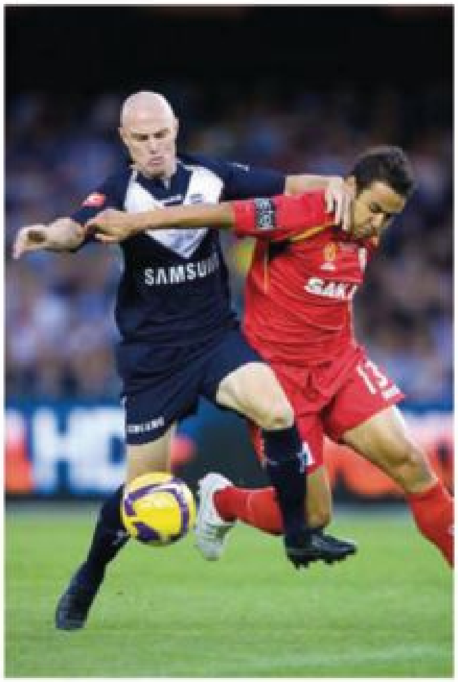


FIGURE 18.1.1 Collisions occur in sport.

FIGURE 18.1.2 A 'bounce' is actually a collision between a ball and the ground. The way a ball bounces is important to many sports.

Collisions occur all the time in sport. Whether it is a ball hitting a bat, a ball bouncing on the ground or two players bumping into each other as shown in Figure 18.1.1, the outcomes of collisions often determine the ultimate result of many sporting contests.

Fortunately, using a couple of relatively simple physics concepts, it is often possible to predict what will happen when two objects collide.

COEFFICIENT OF RESTITUTION

The 'bounce of the ball' plays an important factor in many sports (see Figure 18.1.2). Physicists describe the 'bounciness' of balls using a concept known as the **coefficient of restitution (COR)**. This coefficient is the ratio of the speed of a ball directly after a bounce to its speed before a bounce. As COR is a ratio, it has no units.

$$e = \frac{v_2}{v_1}$$

where e represents the coefficient of restitution

v_1 is the speed before the bounce

v_2 is the speed after the bounce.

The coefficient of restitution actually depends on both the ball and the surface it is bouncing on. A tennis ball bouncing on grass has a different COR than one bouncing on clay. This is one reason why some tennis players prefer to play on some surfaces rather than others.

If there was no energy lost in the bounce, the ball would move just as fast after the bounce as it did before it and e would be equal to 1. However, in reality, some energy is 'lost' in any bounce, so balls generally have a COR of less than 1 (see Figure 18.1.3). The energy is not actually lost, but it has been converted into other forms, such as heat and sound. A coefficient of greater than 1 would mean that the ball had more energy after the bounce than before it, a clear violation of the law of conservation of energy.

Since the COR is defined in terms of speed, v, and kinetic energy, E_k, is proportional to velocity squared, then:

$$e = \frac{v_2}{v_1} = \sqrt{\frac{(E_k)_2}{(E_k)_1}}$$

Consider a ball dropped to the ground (i.e. height = 0) from a height H and rebounding to height h. According to conservation of mechanical energy, the kinetic energy of the ball as it hits the ground is the same as the gravitational potential energy of the ball at the top of the bounce. Therefore,

$$e = \frac{v_2}{v_1} = \sqrt{\frac{(E_k)_2}{(E_k)_1}} = \sqrt{\frac{(E_g)_2}{(E_g)_1}} = \sqrt{\frac{mgh}{mgH}} = \sqrt{\frac{h}{H}}$$

So the coefficient of restitution can be calculated from the initial drop height and the height of the first bounce. This is how many sports define the coefficients of restitution of the balls used to play them.

For example, according to the International Table Tennis Federation rules, a table tennis ball must bounce between 24 and 26 cm when dropped from a height of 30.5 cm onto a steel block. This corresponds to a COR of between 0.89 and 0.92. Similarly, a typical basketball has a value of e in the range of 0.81–0.85, whereas for a tennis ball it is required to be between 0.73 and 0.76.

Worked example 18.1.1

CALCULATING COEFFICIENT OF RESTITUTION

A 'superball' dropped from a height of 100 cm rebounds to a height of 81 cm. Calculate the coefficient of restitution of the ball.	
Thinking	**Working**
Recall the definition of the coefficient of restitution in terms of rebound height.	$e = \sqrt{\frac{h}{H}}$
Calculate the coefficient of restitution.	$e = \sqrt{\frac{81}{100}}$ $= 0.9$

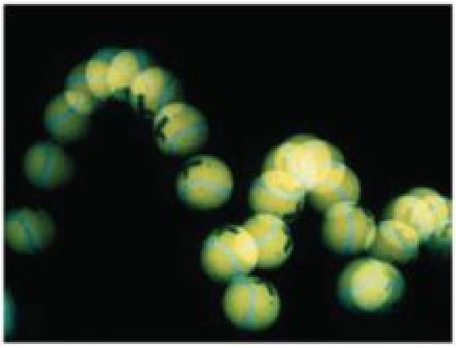


FIGURE 18.1.3 After each bounce, the ball has less energy. The ball has less speed after a bounce and also bounces to a lower height. The COR is less than 1.

Worked example: Try yourself 18.1.1

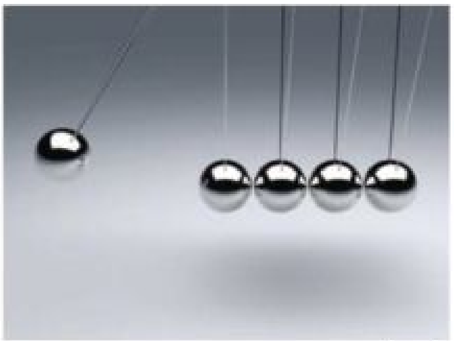
CALCULATING COEFFICIENT OF RESTITUTION

A new tennis ball dropped from shoulder height (1.6 m) rebounds to waist height (1.1 m). Calculate the coefficient of restitution of the tennis ball. Give your answer correct to two significant figures.

TYPES OF COLLISIONS

If an **elastic** material is stretched, compressed or twisted, it returns to its original shape. Traditionally, you might associate the word elastic with elastic bands and springs, but many materials, including most metals, are elastic within certain limits.

Collisions between elastic objects have important properties. Consider a metal ball (ball 1) colliding with four other stationary metal balls in a Newton's cradle as seen in Figure 18.1.4.



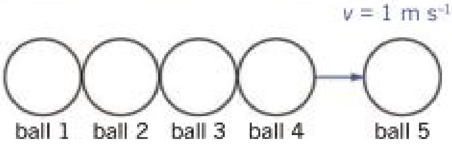


FIGURE 18.1.4 Balls colliding in a Newton's cradle.

The law of conservation of momentum, which was addressed in Chapter 10, applies to any collision, so it can be used to analyse this situation. Assuming that each ball has a mass of 100 g and that the moving ball has a speed of 1 m s^{-1} just prior to the moment of impact, the total initial momentum of the system is:

Total initial momentum of system
= momentum of ball 1 + momentum of balls 2 to 5

$$\Sigma p = mu_1 + mu_{2-5}$$

$$\Sigma p = (0.1 \times 1) + 4 \times (0.1 \times 0) = 0.1 \text{ kg m s}^{-1}$$

By itself, the law of conservation of momentum does not predict what will happen next. In theory, there is an infinite number of ways the balls could move after the collision that would see momentum conserved.

- Scenario A: Ball 1 could stop and ball 5 could gain a velocity of 1 m s^{-1}, as shown in Figure 18.1.5(a) on page 606.

  $$\Sigma p = 4 \times (0.1)(0) + (0.1)(1) = 0.1 \text{ kg m s}^{-1}$$

- Scenario B: All five balls could move together with a velocity of 0.2 m s^{-1}, as shown in Figure 18.1.5(b).

  $$\Sigma p = 5 \times (0.1)(0.2) = 0.1 \text{ kg m s}^{-1}$$

- Scenario C: Ball 1 could rebound at 1 m s^{-1} while balls 4 and 5 gain velocities of 1 m s^{-1}, as shown in Figure 18.1.5(c).

  $$\Sigma p = (0.1)(-1) + 2 \times (0.1)(0) + 2 \times (0.1)(1) = 0.1 \text{ kg m s}^{-1}$$

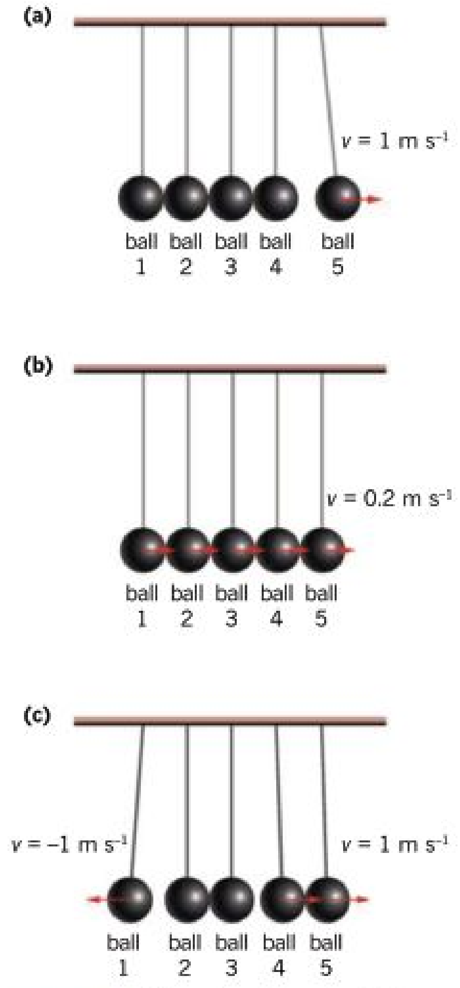


FIGURE 18.1.5 These three different collision outcomes all have the same total momentum.

The key to predicting what will happen in this collision is to consider the amount of kinetic energy in each scenario. Initially, ball 1 has a kinetic energy of $\Sigma E_k = \frac{1}{2}mv^2 = \frac{1}{2}(0.1)(1)^2 = 0.05\text{ J} = 50\text{ mJ}$. Clearly, none of the other balls has any kinetic energy because they are not moving. The law of conservation of energy states that energy cannot be created or destroyed. Therefore, the total energy (in all forms) within the system after the collision must be equal to 50 mJ.

Calculating the total kinetic energy of each scenario gives:

- Scenario A: $\Sigma E_k = 4 \times \frac{1}{2}(0.1)(0)^2 + \frac{1}{2}(0.1)(1)^2 = 0.05\text{ J} = 50\text{ mJ}$
- Scenario B: $\Sigma E_k = 5 \times \frac{1}{2}(0.1)(0.2)^2 = 0.01\text{ J} = 10\text{ mJ}$
- Scenario C: $\Sigma E_k = \frac{1}{2}(0.1)(-1)^2 + 2 \times \frac{1}{2}(0.1)(0)^2 + 2 \times \frac{1}{2}(0.1)(1)^2 = 0.15\text{ J}$
  $= 150\text{ mJ}$

In scenario A, all of the kinetic energy of ball 1 has been effectively transferred directly to ball 5. This type of collision, where kinetic energy has been conserved, is known as an **elastic collision**.

In scenario B, the amount of kinetic energy after the collision is much less than the kinetic energy before the collision. This is possible if some of the original kinetic energy is transformed to other forms of energy such as heat or sound. A collision in which kinetic energy is not conserved is known as an **inelastic collision**.

In scenario C, there is more kinetic energy after the collision than before it. This is an impossible scenario unless there is some source of additional energy that can cause the balls to 'explode' outwards with extra kinetic energy.

Since scenario C is impossible, the outcome must be either scenario A or B. Anyone who has used a Newton's cradle will be able to identify that what actually happens is very close to scenario A, the elastic collision. This is because, in most Newton's cradles, the balls are made from steel or some other elastic material. Because these balls return to their original shape after impact, there is very little energy lost to other forms.

Energy of collisions

An elastic collision is one in which kinetic energy is conserved. That is, the sum of the kinetic energies of all the objects before the collisions is the same as the sum of the kinetic energies after the collision.

Mathematically, this can be written as:

$$\Sigma E_{k\text{ initial}} = \Sigma E_{k\text{ final}}$$

where $\Sigma E_{k\text{ initial}}$ is the total kinetic energy before the collision

$\Sigma E_{k\text{ final}}$ is the total kinetic energy after the collision.

In reality, no macroscopic (large-scale) collisions are perfectly elastic. Even in a Newton's cradle, a small amount of energy is converted into heat and sound. This means that, with each collision, the velocity of the moving ball decreases slightly.

It is possible to get an outcome like scenario B using a Newton's cradle by putting a piece of plasticine onto ball 1 as seen in Figure 18.1.6. The plasticine will absorb much of the kinetic energy of the collision, producing an inelastic collision.

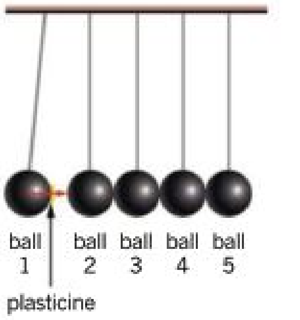


FIGURE 18.1.6 Placing a piece of plasticine on one of the balls will cause the Newton's cradle to produce an inelastic collision.

EXTENSION

Top spin

Although perfectly elastic collisions only occur on the sub-microscopic scale (e.g. between atoms or subatomic particles), some everyday collisions are close enough to being elastic to make their outcomes predictable. For example, a collision between two balls in billiards or pool (see Figure 18.1.7) produces very little heat and sound, and so this collision is close to being elastic.

FIGURE 18.1.7 Pool, billiards and snooker are all games which involve nearly elastic collisions between balls.

This should mean that when a moving billiard ball strikes a stationary billiard ball as shown in Figure 18.1.8, all of the momentum and kinetic energy should be transferred. The moving ball should stop and the stationary ball moves off at close to the initial velocity of the moving ball.

FIGURE 18.1.8 Since collisions between billiard balls are nearly elastic, both momentum and kinetic energy are transferred from the moving ball to the stationary ball.

However, in practice, a much greater variety of outcomes is possible from a billiard ball collision. In fact, skilled billiards players are able to strike the cue ball (the white ball in Figures 18.1.9 and 18.1.10) in ways that can cause it to 'follow' or to 'draw'. To 'follow' means that the moving ball continues rolling in its original direction after the collision. This is shown in Figure 18.1.9. To 'draw' means that the moving ball rolls backwards after the collision. This is shown in Figure 18.1.10.

FIGURE 18.1.9 Billiards players can strike the cue ball so that, after the collision, it 'follows' the ball that it has struck.

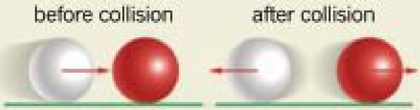


FIGURE 18.1.10 In a 'draw' shot, the player strikes the cue ball so that after the collision, it bounces back in the opposite direction.

These collisions appear to break the law of conservation of energy since kinetic energy seems to have either been created or destroyed. However, this is not the case. These sorts of shots are possible because there is more than one type of kinetic energy involved.

The energy that moves a ball from one spot on the table to another is known as **translational kinetic energy**. However, by hitting the cue ball (the white ball) slightly higher than its centre of mass, as seen in Figure 18.1.11, a billiards player can give the ball some top spin. This gives the ball some **rotational kinetic energy**, or energy of rotation, as well as translational kinetic energy. When the cue ball collides with another ball, some of the rotational kinetic energy is converted into translational kinetic energy, causing the cue ball to follow the ball it has struck.

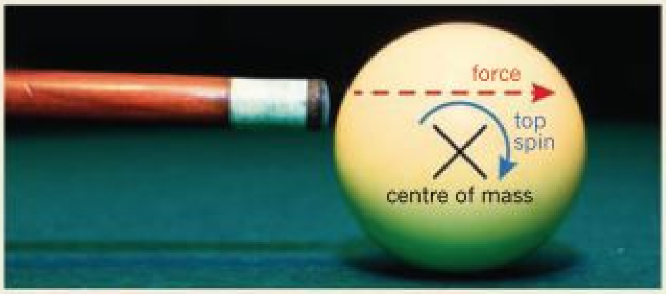


FIGURE 18.1.11 A billiards player can hit the cue ball slightly above the centre of mass of the ball to give it top spin.

Similarly, by striking the cue ball below its centre of mass, the ball can be given back spin. This rotational kinetic energy will cause the ball to move backwards after the collision.

Imparting spin to a ball is a common technique in many sports. The extra rotational energy that this gives the ball is not always obvious to other competitors and it can cause the ball to move or bounce in unexpected ways.

18.1 Review

SUMMARY

- The coefficient of restitution, e, is the ratio of the speed of a ball directly after a bounce to its speed before a bounce: $e = \frac{v_2}{v_1}$
- The coefficient of restitution can be calculated from the rebound height, h, of a ball dropped from height H: $e = \sqrt{\frac{h}{H}}$
- Conservation of momentum can be used to predict the outcome of collisions in sport.
- An elastic collision is one in which kinetic energy is conserved.
- An inelastic collision is one in which kinetic energy is not conserved.
- Momentum is conserved in all collisions.

KEY QUESTIONS

1 All balls have a coefficient of restitution less than 1. This is because

A The atomic bonds between the atoms are broken during collisions.

B The law of conservation of energy prevents it from being any higher.

C All scientific coefficients must be less than 1.

D The ball material is not solid enough to have a better coefficient of restitution than this.

2 Samantha buys a number of new tennis balls whose coefficient of restitution, e, is required to be in the range of 0.73–0.76 for use in tournaments. She drops one of the tennis balls from a height of 70 cm and it rebounds to 45 cm. Samantha decides to reject it. Explain why. Use a calculation to support your answer.

3 A tennis ball of mass 60 g is dropped from a height of 1 m. The ball has a coefficient of restitution of 0.75. Calculate its rebound height. Give your answer correct to two significant figures.

4 A tennis ball is travelling at 60 m s^{-1} when it hits the ground. If it has a coefficient of restitution of 0.82, how much speed will it lose in the bounce?

5 A white billiard ball moving with a speed of 5 m s^{-1} strikes a stationary red ball. After the collision the white ball is stationary and the red ball rolls in the direction of the white ball's initial motion. Each ball has a mass of 160 g.

a Calculate the following:

  i the total initial momentum of the system

  ii the final velocity of the red ball

  iii the initial kinetic energy of the system

  iv the final kinetic energy of the system.

b Is the collision elastic or inelastic?

6 A white billiard ball moving at 4 m s^{-1} collides with a stationary red ball. Assuming both balls have the same mass and the collision is elastic, what would the velocity of the white and red balls be after the collision?

7 A 100 g apple is balanced on the head of a young boy. William, the boy's father, fires an arrow with a mass of 80 g at the apple. It reaches the apple with a velocity of 35 m s^{-1}. The arrow passes right through the apple and goes on with a velocity of 25 m s^{-1}. Assuming there is no friction between the apple and the boy's head, calculate the speed with which the apple will fly off the boy's head.

8 In a Newton's cradle, one ball is dropped so that it has a velocity of 1.5 m s^{-1} as it collides with the other four stationary balls. After the collision, all five balls move together with the same speed. Each ball has a mass of 40 g. Use the law of conservation of momentum to determine:

a the speed of the five balls after the collision

b if the collision is elastic or inelastic.

9 A cricket ball (m = 160 g) reaches a batter at a velocity of 40 m s^{-1}. It is hit by a 1.3 kg bat travelling in exactly the opposite direction at 15 m s^{-1}. If the ball comes off the bat moving at 21 m s^{-1} in exactly the opposite direction to its original direction of travel, determine the following. Give your answers correct to two significant figures.

a the change in momentum of the ball

b the speed of the bat after the collision

c whether the collision is elastic or inelastic

18.2 Sliding and rolling

Many ball sports involve kicking, hitting or throwing balls. In these situations, the size, shape and surface properties of the ball must be considered in order to predict its motion. The force of friction becomes significant, and you must find a way to describe the rotation of the ball along with its translational movement from one point to another.

HOW OBJECTS SLIDE

Friction

When a body slides over a rough surface, friction opposes its motion. The size of the frictional force depends on the nature of the surfaces and the size of the normal force (the normal force was discussed in more detail in Chapter 10). On a horizontal surface, the normal force is equal to the weight of the body, although it acts in the opposite direction (i.e. upwards) as shown by the yellow arrow in Figure 18.2.1.

The relationship between the force of friction, F_f, and the normal force, F_N, is expressed as:

$$F_f = \mu F_N$$

The term μ is the coefficient of friction. The coefficient of friction is a number without units. Its value depends on the type of surface involved and whether it is rough, polished, wet, dry and so on. A low coefficient of friction indicates a small degree of friction.

Frictional forces can be measured in different situations, leading to kinetic friction and static friction.

- **Kinetic friction** (also known as sliding friction) applies when one body is moving across a surface. The coefficient of kinetic friction (μ_k) is used to determine values of kinetic friction.
- **Static friction** is the force that keeps a body stationary even when a pushing or pulling force is acting on the body. The coefficient of static friction (μ_{st}) is used to calculate values of static friction.

Some approximate values of kinetic and static friction for different surfaces are given in Table 18.2.1. Notice that the static friction coefficients are always greater than the kinetic friction coefficients. This is because the forces that act between the surfaces are weakened when the surfaces are moving past each other, and the bonds are not able to form as strongly.

When a ball slides, it experiences kinetic friction because the same part of the ball touches the ground all the time (see Figure 18.2.2).

In comparison, a rolling ball experiences static friction since each section of the ball makes contact with only one section of the ground (see Figure 18.2.3).

Of course, it is not just balls that experience static friction. Consider the friction between a sportsperson's boots and the ground. The static friction between the player's boots and the ground is also important because this is what prevents them from losing traction (grip) and sliding across the ground.

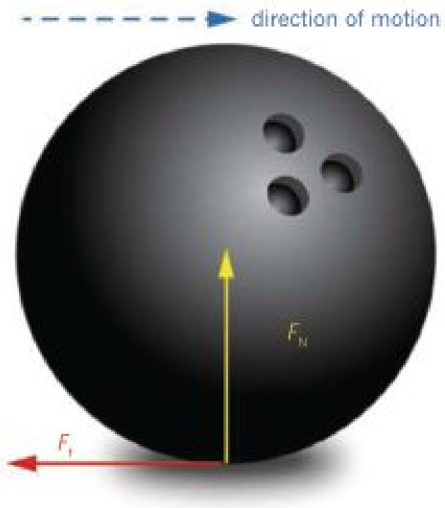


FIGURE 18.2.1 A sliding bowling ball experiences two forces from the ground: the force of friction, F_f, and the normal force, F_N.

Surfaces	μ_{st}	μ_k
wood on wood	0.4	0.2
steel on steel	0.7	0.6
ice on ice	0.1	0.03
steel on ice	0.1	0.03
rubber on dry concrete	0.1	0.8
rubber on wet concrete	0.7	0.5
human joints	0.01	0.01
Teflon on Teflon	0.04	0.04

TABLE 18.2.1 Coefficients of static friction (μ_{st}) and kinetic friction (μ_k) for various combinations of surfaces.

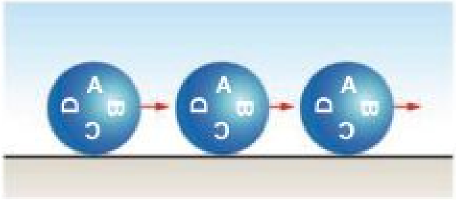


FIGURE 18.2.2 A sliding ball undergoes translation motion. The same section of the ball (C) is in contact with the bowling alley at all times.

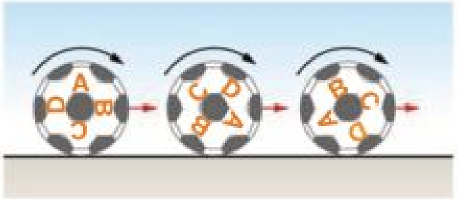


FIGURE 18.2.3 For a rolling ball, the surface of the ball and the ground are stationary relative to each other while they are in contact.

Worked example 18.2.1

CALCULATING FRICTION

The coefficient of kinetic friction between a 7.0 kg bowling ball and the lane is 0.08. Calculate the force of friction acting on the ball.	
Thinking	**Working**
Calculate the normal force acting on the ball. On a flat surface, the normal force will be equal in magnitude (but opposite in direction) to the weight of the ball.	$F_N = -F_g = -mg = -7.0 \times 9.8$ $= -69$ N (The negative sign indicates that the normal force is in the opposite direction to gravity.)
Recall the formula for friction.	$F_f = \mu F_N$
Calculate the force of friction.	$F_f = 0.08 \times -69$ $= -5.5$ N (The negative sign indicates that friction is a retarding force.)

Worked example: Try yourself 18.2.1

CALCULATING FRICTION

Calculate the force of friction acting on a 7.2 kg bowling ball if the coefficient of kinetic friction between the ball and the lane is 0.12. Give your answers correct to two significant figures.

Using your knowledge of Newton's laws and the constant acceleration formulas, it is possible to further analyse the motion of sliding and rolling objects. The equations related to constant acceleration were discussed in detail in Chapter 9, and Newton's laws were discussed in Chapter 10.

Worked example 18.2.2

EFFECT OF FRICTION

A 7.2 kg bowling ball slides the entire length of the bowling lane (i.e. 18 m), experiencing a coefficient of kinetic friction of 0.04. If it is released with a speed of 9.0 m s^{-1}, how fast will it be going when it hits the pins? Give your answer correct to two significant figures.	
Thinking	**Working**
Calculate the normal force acting on the ball. On a flat surface, the normal force will be equal in magnitude (but opposite in direction) to the weight of the ball.	$F_N = -F_g = -mg = -7.2 \times 9.8$ $= -71$ N
Recall the formula for friction.	$F_f = \mu F_N$
Calculate the force of friction.	$F_f = 0.04 \times -71$ $= -2.8$ N
Calculate the deceleration of the bowling ball using Newton's second law.	$F = ma$ $\therefore a = \frac{F}{m} = \frac{-2.8}{7.2}$ $a = -0.4$ m s^{-2}
Use an appropriate equation of motion to calculate the final speed (v) of the ball given: $u = 9.0$ m s^{-2} $a = -0.4$ m s^{-2} $s = 18$ m	$v^2 = u^2 + 2as$ $\therefore v^2 = (9.0)^2 + 2 \times -0.4 \times 18$ $v^2 = 67$ $v = \sqrt{67}$ $= 8.2$ m s^{-1}

Worked example: Try yourself 18.2.2

EFFECT OF FRICTION

A 65 kg footballer wears boots that have a coefficient of kinetic friction with grass of 0.69. If the footballer is running at 8.0 m s^{-1} and then suddenly tries to stop, how fast will she be going after sliding for 2 m? Give your answers correct to two significant figures.

THE DIFFERENCE BETWEEN ROLLING AND SLIDING

When analysing the motion of a ball, it is often convenient to treat it as a **point mass**, i.e. a particle with mass but no other dimensions. For example, when considering a tennis ball as a projectile, you can assume that all of the mass of the tennis ball is concentrated at a single point at the centre of the ball (see Figure 18.2.4). This greatly simplifies the situation, since you only have to consider its **translational motion**, that is, its motion as it 'translates' or moves from one point in space to another.

In reality, many tennis players try to hit the ball so that it spins, as in Figure 18.2.5, in a way that makes it more difficult for their opponent to hit. In order to analyse this kind of motion, you have to consider the ball's **rotational motion**.

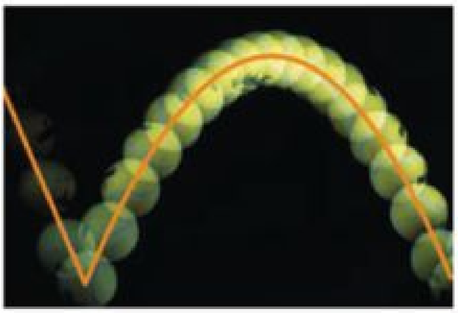


FIGURE 18.2.4 A tennis ball can be treated as if all of its mass is concentrated at its centre. The position of the centre of mass is represented by the orange line.

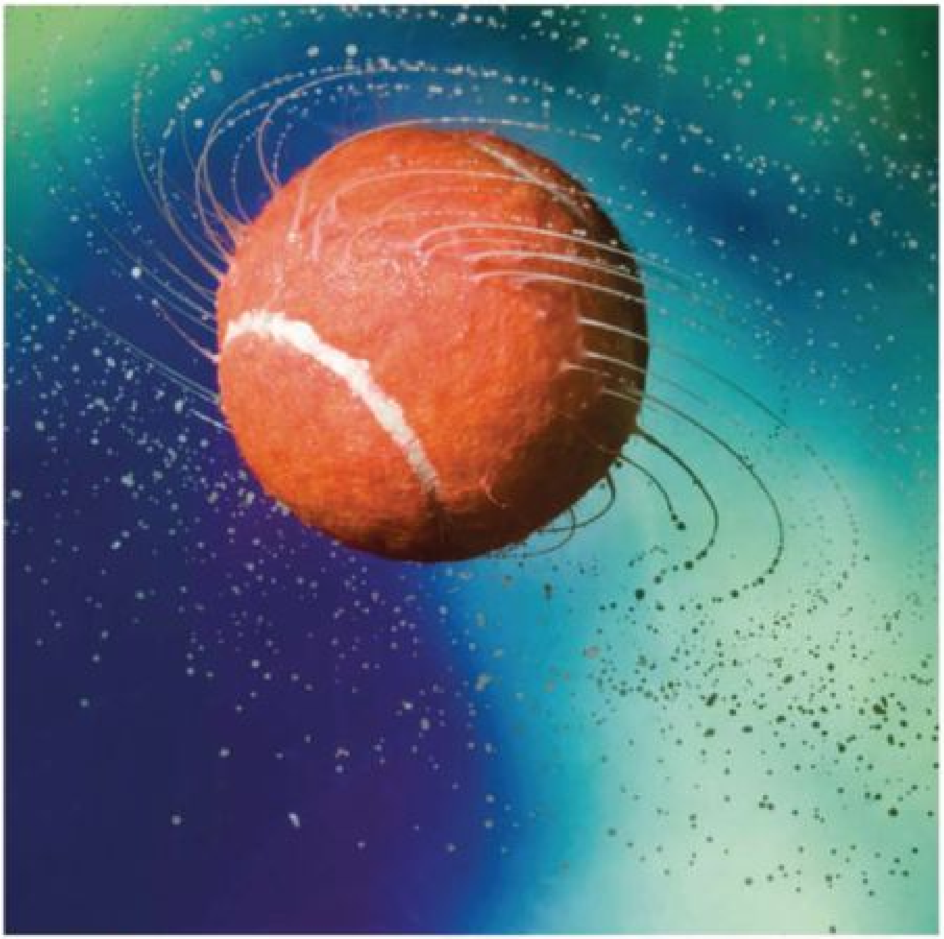


FIGURE 18.2.5 A high-speed photograph of a wet tennis ball shows that it is spinning.

FIGURE 18.2.6 A bowling ball slides for the first few metres down a bowling lane before it starts to spin.

When a ball undergoes translational motion along a surface without any rotational motion, it 'slides'. For example, the wooden floorboards of a bowling alley are polished so that a bowling ball will slide for the first few metres before it starts to spin, as shown in Figure 18.2.6.

In comparison, a ball kicked gently along the ground will roll, which means it undergoes translational motion and rotational motion simultaneously (see Figure 18.2.7).

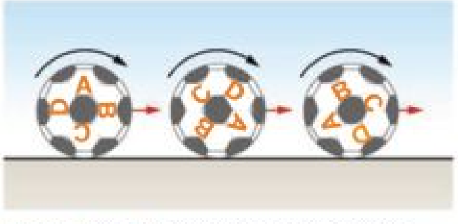


FIGURE 18.2.7 The rotational motion of this ball is shown by the black arrows and the translational motion by the red arrows.

MEASURING ROTATIONAL MOTION

Using radians

Angles are important when measuring rotational motion. In mathematics, you have probably become used to measuring angles in terms of degrees, where one full rotation equals 360°. Although this is a relatively simple and familiar unit of measurement, physicists prefer to measure angles using a unit known as a **radian**.

The mathematical definition of a radian is a little complex: *the measure of a central angle subtending an arc equal in length to the radius*. This essentially means that a radian is defined such that 1 revolution is equal to exactly 2π radians.

1 revolution = 2π radians

So, 360° = 2π radians and 180° = π radians

Therefore:

2π radians = 360°

$\therefore \pi$ radians = 180°

$\therefore 1$ radian $= \frac{180^\circ}{\pi} \approx 57.3^\circ$.

A visual representation of 1 radian is shown in Figure 18.2.8

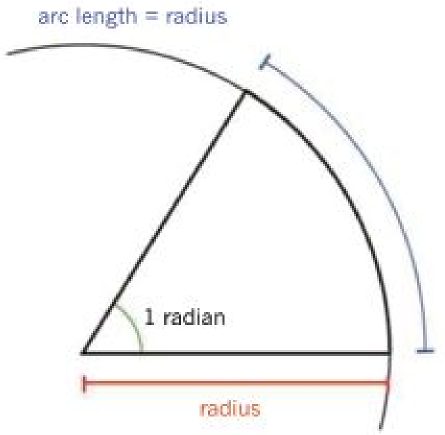


FIGURE 18.2.8 A radian is defined as the angle corresponding to an arc equal to the radius of the circle on its circumference.

Since π is an irrational number (i.e. it cannot be expressed as an exact ratio), wherever possible, mathematicians prefer to express angles measured in radians as common fractions or multiples of π. However, in science, angles often have to be expressed as decimal fractions. Some common angles and their decimal values are expressed in Table 18.2.2.

Angle (in degrees)	Exact angle (in radians)	Approximate angle (in radians)
30°	$\frac{\pi}{6}$	0.524
45°	$\frac{\pi}{4}$	0.785
60°	$\frac{\pi}{3}$	1.047
90°	$\frac{\pi}{2}$	1.571
120°	$\frac{2\pi}{3}$	2.094
135°	$\frac{3\pi}{4}$	2.356
180°	π	3.142
360°	2π	6.283

TABLE 18.2.2 Common angles in degrees and radians.

Worked example 18.2.3

CONVERTING BETWEEN DEGREES AND RADIANS

Make the following conversions. Give your answers correct to three significant figures.

a 24° to radians

Thinking	Working
Recall the conversion for degrees to radians.	Multiply value by $\frac{\pi}{180}$
Convert 24° to radians.	$24 \times \frac{\pi}{180} \approx 0.419$

b 1.25 radians to degrees

Thinking	Working
Recall the conversion for radians to degrees.	Multiply value by $\frac{180}{\pi}$
Convert 1.25 radians to degrees.	$24 \times \frac{180}{\pi} \approx 71.6^\circ$

Worked example: Try yourself 18.2.3

CONVERTING BETWEEN DEGREES AND RADIANS

Make the following conversions. Give your answers correct to three significant figures.

a 130° to radians

b 5.65 radians to degrees

PHYSICSFILE

Degrees and radians

Converting from degrees to radians
Multiply value by $\frac{\pi}{180}$
Converting from radians to degrees
Multiply value by $\frac{180}{\pi}$

TABLE 18.2.3 Converting degrees and radians.

Angular speed and linear speed

Just as you can measure the speed of a ball as it moves in a straight line from one point to another (i.e. its 'linear' speed), you can also measure its speed of rotation. This is known as the ball's **angular speed**.

Angular speed is defined as:

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

where ω is angular speed

θ is the angle

t is time.

If time is measured in seconds (s) and angle is measured in radians (rad), then the units for angular speed are rad s^{-1}.

Sometimes balls can spin very rapidly, in which case it is easier to describe their frequency of rotation. Frequency (f) is measured in Hertz (Hz) where 1 Hz = 1 rotation per second.

Frequency can be easily converted into a time of rotation using the formula

For a full circle, $\theta = 2\pi$ rad, therefore

$$\omega = \frac{2\pi}{t} = 2\pi f$$

where f is the frequency (in Hz or rotations per second)

Angular velocity as a vector quantity

Combining an object's linear speed with its direction of motion gives its linear velocity. That is, an object's linear velocity has a magnitude (speed) and a direction as it is a vector. Similarly, the angular velocity of an object is a vector quantity that describes its angular speed and direction of rotation. It can be complicated to mathematically represent the direction of rotation of an object; however, it can usually be relatively simply expressed as either *clockwise* or *anticlockwise* relative to the observer.

Worked example 18.2.4

CALCULATING ANGULAR SPEED

A spinning cricket ball goes through 20 complete rotations each second. Calculate its angular speed. Give your answers correct to three significant figures.

Thinking	Working
Convert the frequency to a time of rotation.	$t = \frac{1}{f} = \frac{1}{20}$ $= 0.05$ s
Recall that 1 rotation = 2π radians.	$\theta = 2\pi \approx 6.238$ rad
Recall the definition of angular speed.	$\omega = \frac{\theta}{t}$
Calculate the angular speed. (Note that the rule $\omega = 2\pi f$ could also have been used.)	$\omega = \frac{6.283}{0.05}$ $= 126$ rad s^{-1}

Worked example: Try yourself 18.2.4

CALCULATING ANGULAR SPEED

A spinning tennis ball goes through 50 complete rotations in 1 second. Calculate its angular speed. Give your answers correct to three significant figures.

Converting between angular speed and linear speed

The linear speed of a rolling ball can be calculated from its angular speed if the radius of the ball is known. The formula to convert angular speed to linear speed is:

$v = r\omega$

where v is linear speed (in m s^{-1})

r is the radius of the ball (in m)

ω is the angular speed (in rad s^{-1}).

This rule is based on the idea that for each full revolution, a ball will roll a distance equal to its circumference, as shown in Figure 18.2.9.

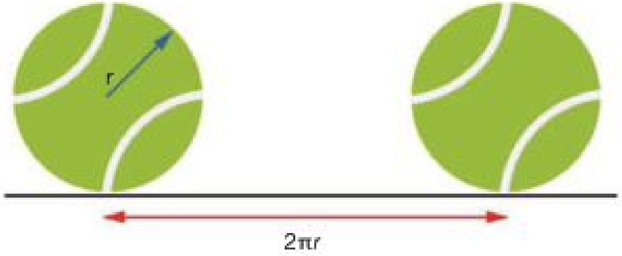


FIGURE 18.2.9 The roll distance is equal to the circumference for a full revolution of a ball.

The use of radians to measure angles makes conversions between angular and linear speeds much easier to calculate. If the angular speed, ω, was measured in degrees per second rather than radians per second, then the conversion formula would be:

$$v = \frac{\omega}{360} \times 2\pi r$$

Worked example 18.2.5

CONVERTING ANGULAR SPEED TO LINEAR SPEED

A tennis ball with a diameter of 6.7 cm rolls along the ground at an angular speed of 10 rad s^{-1}. Calculate its linear speed.

Thinking	Working
Calculate the radius of the tennis ball in m.	$r = \frac{1}{2}D = \frac{1}{2} \times 0.067$ $= 0.0335$ m
Recall the formula for converting angular speed to linear speed.	$v = r\omega$
Calculate the linear speed.	$v = 0.0335 \times 10$ $= 0.335$ m s^{-1}

Worked example: Try yourself 18.2.5

CONVERTING ANGULAR SPEED TO LINEAR SPEED

A basketball with a diameter of 75 cm rolls along the ground at an angular speed of 40 rad s^{-1}. Calculate its linear speed.

18.2 Review

SUMMARY

- The relationship between the force of friction, F_f, and the normal force, F_N, is expressed as:
  $F_f = \mu F_N$
  where μ is the coefficient of friction.
- A stationary object experiences static friction; a sliding object experiences kinetic friction.
- The coefficient of kinetic friction (μ_k) is used to calculate values of kinetic friction.
- The coefficient of static friction (μ_{st}) is used to calculate values of static friction.
- The motion of a rolling ball can be described using either angular speed, ω, or linear speed, v, where $v = r\omega$.
- Angular motion is measured in radians, where π radians = 180°

KEY QUESTIONS

1 Scientific studies show that a static friction of 0.8 between an athlete's boots and the ground is sufficient for most sports. Based on this figure, how much friction is required for an athlete with a mass of 75 kg? Use $g = 9.8\ \text{m s}^{-2}$.

A 60 N
B 94 N
C 590 N
D 920 N

2 When a ball rolls along the ground it experiences which type of friction?

A zero
B kinetic
C static

3 An ice hockey puck (m = 165 g) is hit with an initial speed of 41 m s^{-1} on a large frozen lake. If the coefficient of kinetic friction between the puck and the ice is 0.038, what is the speed of the puck after 7.0 s? Give your answer correct to two significant figures.

4 An 80 kg baseball player begins a slide into third base when moving at a speed of 5.0 m s^{-1}. The coefficient of friction between his body and the dirt is 0.65. How far does he slide? Give your answer correct to two significant figures and use $g = 9.8\ \text{m s}^{-2}$.

5 A soccer ball is kicked so that it skids along the ground for a distance of 25 m before coming to a stop. If the initial speed of the ball is 15 m s^{-1} and the ball has a mass of 440 g, calculate the average coefficient of friction experienced by the ball. Give your answer correct to two significant figures and use $g = 9.8\ \text{m s}^{-2}$.

6 By definition, the number of radians in a whole circle is:

A 3.14
B π
C 2π
D 360π

7 Copy and complete the table below by filling in the equivalent angle. State angles to the nearest whole degree or hundredth of a radian. Give your answers correct to three significant figures.

Angle (in degrees)	Angle (in radians)
57°	
125°	
280°	
450°	
	0.50
	1.25
	4.80
	7.12

8 Use the definition of angular speed to fill in the blanks in the table below. Round your answers to two significant figures where necessary.

Angle, θ (rad)	Time, t (s)	Angular speed ω (rad s^{-1})
3.14	0.01	
48	0.3	
630	21	
	2.4	25
	0.34	54
	0.1	8.8
55		6.0
8.4		38

9 A bowling ball has a diameter of 21.7 cm and an angular speed of 85.7 rad s^{-1}. Calculate its linear speed.

10 A bowling alley is 18.3 m long. If it takes 2.50 s for a bowling ball with a diameter of 21.6 cm to roll the length of the bowling alley, calculate the average linear and angular speeds of the ball. Give your answers correct to three significant figures.

18.3 Hitting, kicking or throwing

The process of hitting, kicking or throwing a ball may seem simple, but on closer investigation there's much more to it. What seems like a simple swing of the racquet in Figure 18.3.1 actually involves complex energy transfers and transformations. These processes can be understood using physical models such as the simple **pendulum**.

FIGURE 18.3.1 Bats and racquets are involved in the process of energy transfer.

SIMPLE PENDULUM

A simple pendulum consists of a mass called a bob, suspended by a length of rope or string from a fixed point. When the mass is moved slightly to one side and released, it swings back and forth as shown in Figure 18.3.2. This type of repeated movement is called **simple harmonic motion**.

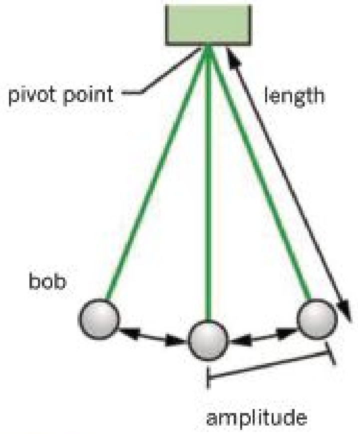


FIGURE 18.3.2 A simple pendulum showing its length, amplitude and the path taken by the bob (mass) for one oscillation.

ENERGY CHANGES IN A SIMPLE PENDULUM

A swinging pendulum provides an example of the conservation of mechanical energy.

Consider the simple pendulum shown in Figure 18.3.3.

- If you take the lowest point in the bob's swing as the zero potential energy level, then at the top of its swing, the bob has its maximum gravitational potential energy, $E_g = mg\Delta h$.
- At the bottom of the bob's swing, all of this gravitational potential energy will be converted into kinetic energy ($E_k = \frac{1}{2}mv^2$). This is the point in the swing when the bob has its maximum velocity.
- As it swings through to the other side, the kinetic energy of the bob is converted back into potential energy.

In a perfectly frictionless pendulum, the bob would reach the same height on either side of the pendulum. In reality, each swing becomes successively smaller, as energy is lost as heat due to friction.

Conservation of mechanical energy explains why a cricket player will use a high backswing to hit the ball to the boundary fence (see Figure 18.3.4). Lifting the bat to head height, gives it a large amount of gravitational potential energy. When the bat swings down, this potential energy is converted into kinetic energy that is transferred to the ball.

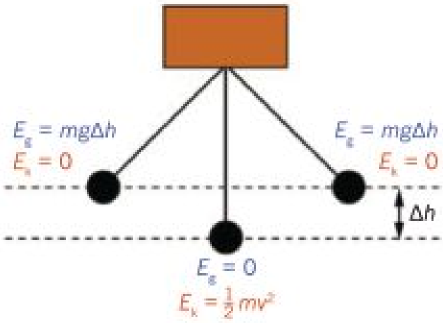


FIGURE 18.3.3 In a simple pendulum, energy is transformed from gravitational potential energy to kinetic and energy and back.

FIGURE 18.3.4 A high backswing means the bat has high gravitational potential energy. This will be converted to high kinetic energy when the bat hits the ball.

In this situation, the bat is acting as a *driven pendulum*. So in addition to the force of gravity, the batter also applies a force to the bat that gives it even more kinetic energy when it hits the ball.

When the bat hits the ball, some of its kinetic energy is transferred to the ball. This causes the ball to move in a different direction. The bat also retains some kinetic energy, which is why batters need to 'follow through' in their shots as shown in Figure 18.3.5.

FIGURE 18.3.5 A batter follows through the shot because the bat retains some kinetic energy after it hits the ball.

Some of the kinetic energy of the bat is also transformed into the sound of the bat hitting the ball and into heat energy. This loss of energy means that the collision between a bat and ball is an inelastic collision (see Section 18.1).

PHYSICSFILE

Hot spots

When a cricket ball hits a cricket bat, there is a small amount of friction between the ball and the bat. This friction results in some of the energy of the ball being converted to heat at the point of contact with the bat. This is the basis of the 'Hot Spot' technology which can be used to detect whether or not the ball has hit the bat. 'Hot Spot' uses an infrared camera to show the increased temperature at the point of contact as a white spot.

THE PERIOD OF A SIMPLE PENDULUM

The period of the pendulum is the time taken for one complete swing or oscillation. A complete oscillation occurs when the mass swings from one side to another and back again, returning to its original position.

The period of a pendulum (T) is given by the formula:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

where T is the period of oscillation (in s)

L is the length of the pendulum (in m)

g is the acceleration to due gravity (9.8 m s^{-2}).

The equation above shows that the period of the pendulum depends only on its length. Changing the mass of the bob or the amplitude of the swing has no effect on it. This equation is only true as long as the mass of the string or rope is small and the amplitude of the swing is small (i.e. the angular displacement of the mass from the mean position is less than or equal to 15°).

Worked example 18.3.1

SIMPLE PENDULUM

Calculate the period of a simple pendulum that is 30 cm long. Use $g = 9.8$ m s^{-2}.	
Thinking	**Working**
Recall the equation for the period of a simple pendulum.	$T = 2\pi\sqrt{\frac{L}{g}}$
Substitute the correct values into this equation.	$L = 0.3$ m and $g = 9.8$ m s^{-2}, therefore: $T = 2 \times 3.14 \times \sqrt{\frac{0.3}{9.8}}$ $= 1.1$ s

Worked example: Try yourself 18.3.1

SIMPLE PENDULUM

Calculate the period of a simple pendulum that is 70 cm long. Give your answer correct to two significant figures and use $g = 9.8$ m s^{-2}.

PHYSICS IN ACTION

Models and theories

Galileo is often identified as one of history's first experimental scientists. He is famous for his discovery that all objects fall at the same rate, his development of the telescope and his historically significant work in astronomy.

However, Galileo's first scientific discovery came from watching a lamp swinging in a cathedral in Pisa (see Figure 18.3.7). He realised that the time taken for one swing depended on the length of the chain holding the lamp. Galileo used this to discover a relationship between length and period for all pendulums.

FIGURE 18.3.7 Galileo developed a hypothesis about the period of a pendulum after watching a lamp swinging in a cathedral in Pisa.

In the early seventeenth century, there were very few mechanisms that could be used to keep time reliably. Galileo designed a pendulum clock based on his discovery (see Figure 18.3.8).

FIGURE 18.3.8 A model of a pendulum clock designed by Galileo.

To this day, this timing mechanism is still used in grandfather clocks as seen in Figure 18.3.9.

FIGURE 18.3.9 A grandfather clock.

DOUBLE PENDULUM

Most biomechanical systems are much more complicated than a simple pendulum. Since your arms and legs are jointed, they can be represented using a *double pendulum*. A double pendulum consists of one pendulum suspended from another pendulum as shown in Figure 18.3.10.

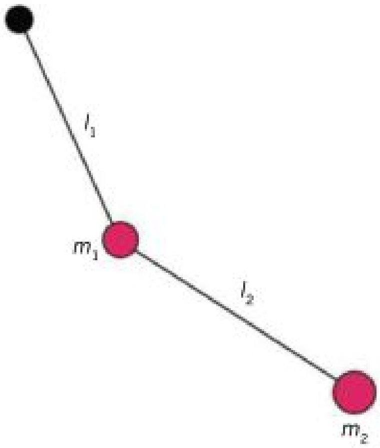


FIGURE 18.3.10 A double pendulum consists of a mass (m_2) suspended from another suspended mass (m_1).

As you can see in Figure 18.3.11, a human arm or leg can be treated as a double pendulum with the foot/hand and lower limb acting as the lower pendulum and the knee/elbow and upper limb acting as the upper pendulum.

FIGURE 18.3.11 A human leg can be modelled using a double pendulum.

In a double pendulum model of a golf swing, the human arms acts as the upper pendulum while the golf club acts as the lower pendulum, as shown in Figure 18.3.12. When a golfer swings at the ball, the goal is to accelerate the club head so that it strikes the ball at just the right point and in the right direction, moving as quickly as possible. To do this, the golfer exerts force with his or her arms on the shaft of the club, which in turn exerts force on the club head.

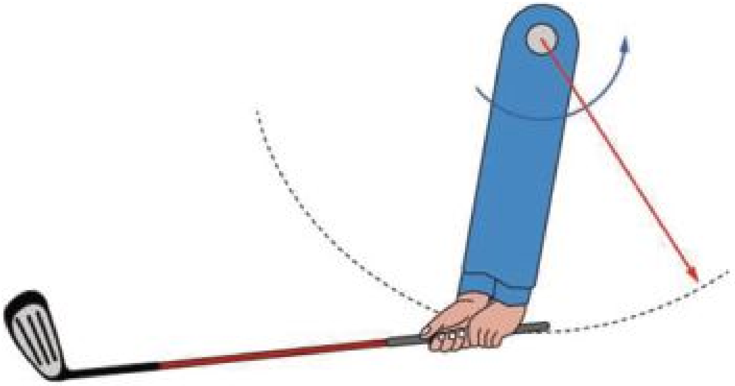


FIGURE 18.3.12 A golf swing can be treated as a double pendulum with the arms acting as one pendulum and the club acting as the other pendulum.

The velocity of the club head, together with its mass, determine its kinetic energy and momentum. At the top of the swing both parts of the pendulum (upper and lower) have potential energy. Looking at Figure 18.3.13, as the swing progresses, the arms move and the golfer applies more and more force to the club head causing it to accelerate and so increase its speed. Therefore the momentum and kinetic energy of the club head increase.

FIGURE 18.3.13 The double pendulum model of a simple golf swing showing upper (yellow) and lower (red) pendulums.

At the point of impact, most of the kinetic energy from the golfer's arms and body have been transferred to club and some of this energy and momentum is transferred to the ball. To determine the speed of the ball as it leaves the tee, you need to consider the conservation of both energy and momentum. In reality, not all of the kinetic energy lost by the club head during impact is converted into kinetic energy of the ball. That is, the impact is not perfectly elastic. Some energy is lost to sound, heat and damage to the ball.

A double pendulum has a much wider range of motion than a simple pendulum. This motion is also much more complicated to analyse because, in addition to gravity, each bob experiences a force from the other bob. In the human body, the forces applied by skeletal muscles also need to be considered.

EXTENSION

Chaotic motion

Scientists use the term *chaotic* to describe systems like the double pendulum. Although seemingly unpredictable, chaotic systems are governed by the same predictable laws of physics as other systems. However, chaotic systems are very sensitive to changes in their initial conditions.

For example, a simple pendulum is not a chaotic system because, regardless of the initial amplitude of the pendulum, the bob will always do the same thing—swing backwards and forwards in simple harmonic motion. In comparison, the outcome of a double pendulum, as seen in Figure 18.3.14, will vary greatly depending on the exact starting position of each of the bobs. Even a small change in the initial position of one bob will mean that the double pendulum system will swing a very different way.

For the same reasons, the transfer of energy during a golf shot is very hard to replicate since the double pendulum undergoes chaotic motion, and shows sensitive dependence on initial conditions. Factors such as how far the club head is taken back and the position of the arms and wrists all contribute to the variety of shots seen on golf courses.

Chaotic systems are common in many other areas of science including meteorology (the study of the weather), microbiology and the study of population dynamics.

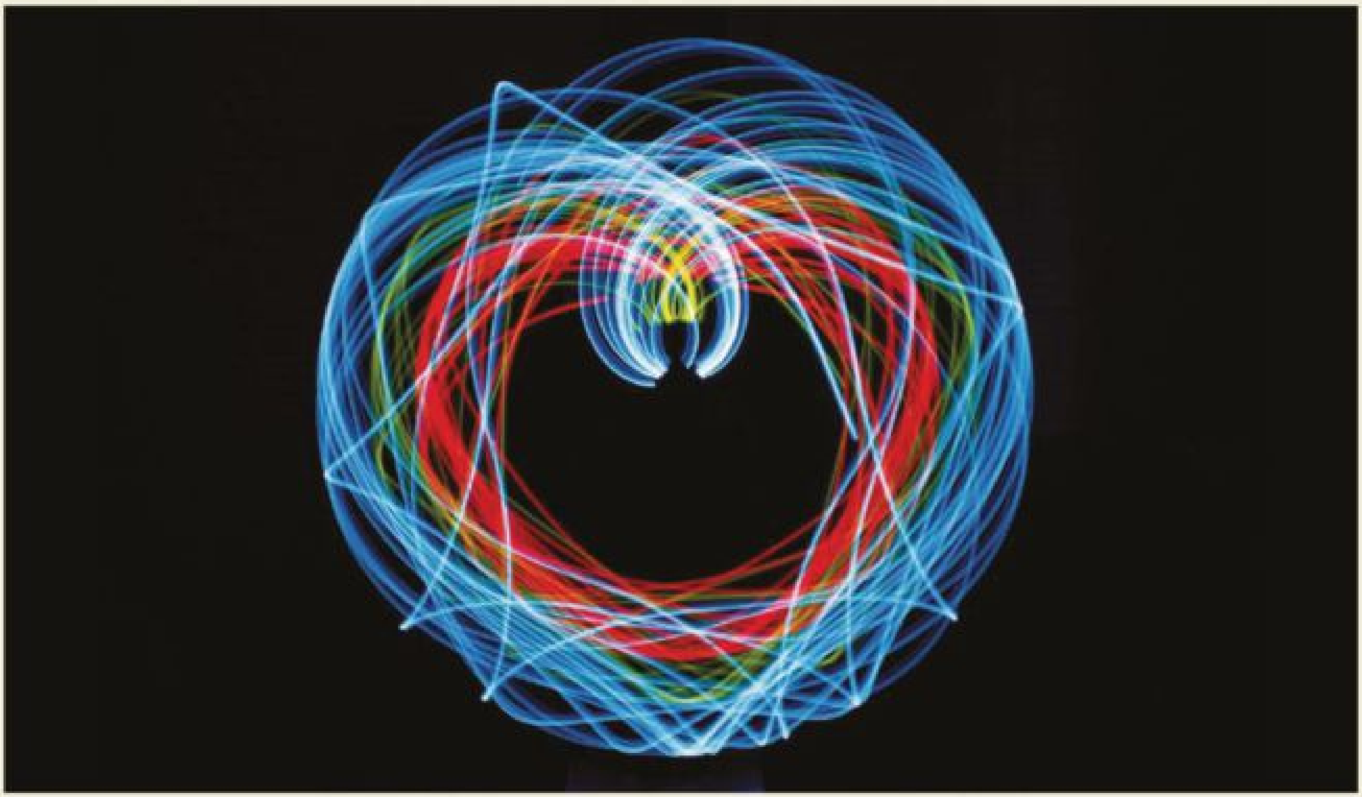


FIGURE 18.3.14 A long-exposure photograph of a double pendulum with a light attached to its second bob demonstrates the range of motion it can undergo.

18.3 Review

SUMMARY

- Biomechanical systems can be modelled using simple or double pendulums.
- The conservation of mechanical energy can be used to understand the energy changes that occur in a pendulum.

KEY QUESTIONS

1 A golfer lines his club up with the ball at the start of his swing as shown. Identify the energy state of the head of the golf club at this point in the swing.

A gravitational potential energy is maximum, kinetic energy is maximum
B gravitational potential energy is zero, kinetic energy is maximum
C gravitational potential energy is maximum, kinetic energy is zero
D gravitational potential energy is zero, kinetic energy is zero

2 The golfer is now at the top of his swing. Explain the energy state of the head of the golf club at this point in the swing. Assume mechanical energy is conserved throughout the swing.

3 When a bat hits a ball, the kinetic energy of the bat:
A is retained by the bat
B is transferred to the ball
C is transformed into heat and sound
D all of the above

4 A simple pendulum consisting of a 150 g bob swings back and forth. What is the amount of energy stored in the bob when it swings 22.5 cm above the lowest point? Give your answer correct to two significant figures and use $g = 9.8\ \text{m s}^{-2}$.

5 Calculate the period of a 30 cm pendulum on the Moon where $g = 1.6\ \text{m s}^{-2}$. Give your answer correct to two significant figures.

6 Calculate the length of a pendulum that would have a period of exactly 1 second. Give your answer correct to the nearest mm.

7 Which of the following would increase the period of a simple pendulum?
A increasing its length
B decreasing its length
C increasing its mass
D decreasing its mass

8 Explain why the footballer shown below has swung her leg so far back before kicking the ball.

9 The picture below shows a person about to kick a football. Sketch a simple diagram and circle each of the individual pendulums that make up the double pendulum in this situation.

10 Explain why a double pendulum is an example of a chaotic system.

18.4 The flight of a ball

Rarely does a ball travel in a straight line on the sports field. Whether it be the spin on the ball caused by a racquet, the dimples of a golf ball breaking air flow or the wind catching a netball and pulling it off course, the flight of any two balls is rarely the same. The path of the ball can be difficult to predict, even for the experienced sportsperson. However, there are some things that every sports ball will have in common while in flight. Once a theoretical understanding of the ideal flight of a ball is understood, the effects of air resistance, drag and spin can be investigated.

In this section, you will apply the straight line equations of motion and your understanding of acceleration under the force of gravity to investigate the flight of balls on a sports field.

FIGURE 18.4.1 A multi-frame image of a tennis ball bouncing. The changes in motion as it bounces are visible through the different distances between the ball in each frame.

INVESTIGATING MOVEMENT IN TWO DIMENSIONS

Analysing the flight of a ball, or any object in flight, involves considering the movement of the object in both horizontal and vertical directions. For example, consider the bouncing ball in Figure 18.4.1. The ball is bouncing up and down as well as moving sideways.

Figure 18.4.2 shows a single bounce of a ball across a smooth flat surface. The circles represent the position of the ball at equal time intervals during the bounce. The arrows at each position give an indication of how the vertical and horizontal position of the ball changes at each point throughout the bounce. So, for example, the biggest change in vertical position is near the start and end of the bounce.

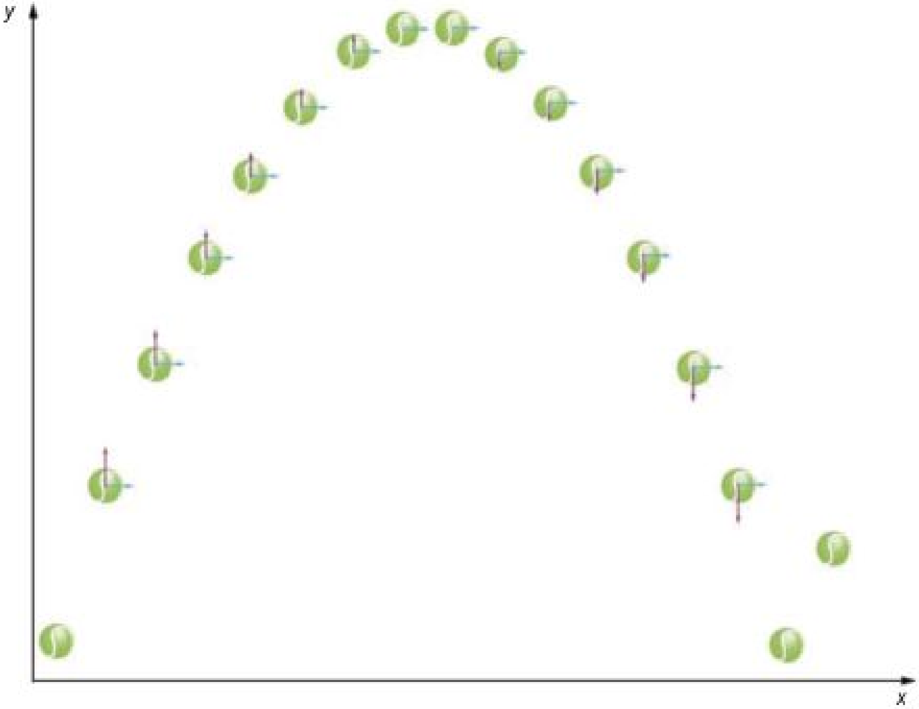


FIGURE 18.4.2 The path of a bouncing ball, from left to right, with arrows showing the changes in its vertical and horizontal position.

The bounce shown in Figure 18.4.2 can be further analysed by plotting the horizontal and vertical positions of the ball on separate graphs. The change in the ball's position from the start of its motion (the origin) is shown in Figure 18.4.3. Graph (a) shows the horizontal displacement and graph (b) the vertical displacement from the origin. Note that the origin of the graph has been established on the left of the field of view. As the ball bounces towards the right, the horizontal position will increase as the ball moves further from the origin. Height on the graph will increase as the ball moves up and will decrease as it moves down.

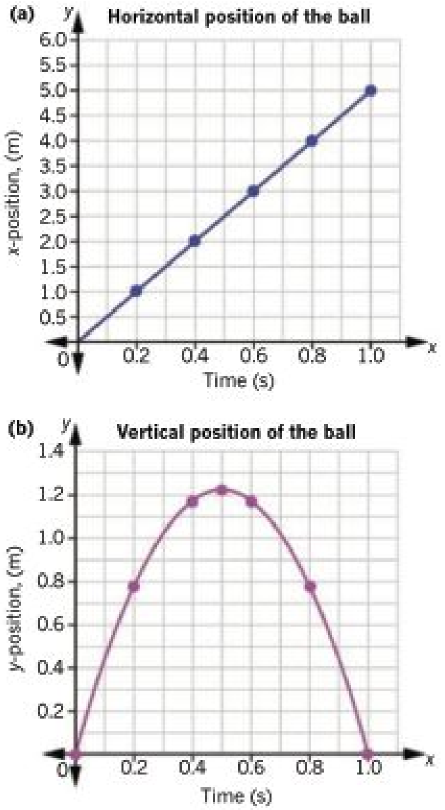


FIGURE 18.4.3 Graphs of the horizontal and vertical displacement of the ball from the origin.

The graph of horizontal position–time (Figure 18.4.3(a)) shows that:

- The distance from the origin increases as the ball moves away from the origin.
- The graph is a straight line with a regular interval between each point.

The graph of vertical position–time (Figure 18.4.3(b)), is quite different.

- The distance from the origin increases as the ball bounces higher, and then decreases as the ball returns to the ground.
- The graph is roughly parabolic in shape with a changing interval between each point. That is, the points on the graph are closer together near the top of the bounce.

The shape of each position–time graph illustrates the following:

- In the horizontal direction, the straight line position–time graph indicates that the velocity of the ball is constant. There is no acceleration in the horizontal direction.
- In the vertical direction, the changing gradient of the position–time graph indicates that the velocity is changing. There is a force affecting the ball and the ball is accelerating.

Velocity–time graphs of horizontal and vertical motion can be used to confirm these conclusions and determine the size and direction of the acceleration found to be affecting the ball in the vertical direction. These are shown in Figure 18.4.4.

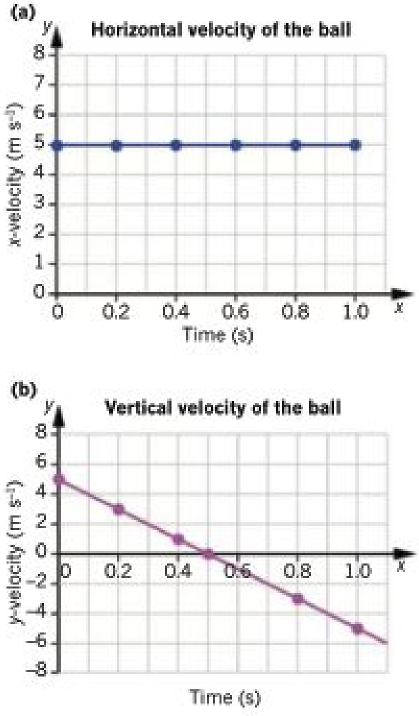


FIGURE 18.4.4 Graphs showing the velocity of the ball during a single bounce. (a) The horizontal velocity is constant. (b) The vertical velocity is changing at a constant rate, implying a constant acceleration.

The graph of horizontal velocity versus time shown in Figure 18.4.4(a) confirms that the horizontal velocity is constant. Once the ball leaves the person's hands, and ignoring air resistance, there are no forces acting in a horizontal direction. The ball will continue to move at the same horizontal velocity as when it was first thrown. The positive value confirms that the ball is moving further from the origin.

In the vertical direction, the gradient of the velocity–time graph confirms that there is an acceleration, and hence a force, affecting the movement of the ball. The constant gradient of the graph (allowing for small measuring errors) indicates that this acceleration is constant. The negative slope indicates that the acceleration is acting downwards. The graph also shows that the ball had a maximum velocity at the bottom of the motion, that is, at the start of the bounce and at the very end of the bounce. Finally, you can see that the velocity was at a minimum (zero) when at the maximum height of the bounce.

Ignoring air resistance, the force that causes the vertical velocity of the ball to change is gravity. Gravity is constant and acts downwards towards the ground, which matches the shape of the velocity–time graph in Figure 18.4.4(b).

So in summary, and ignoring air resistance, a ball or object in flight will:

- continue at a constant velocity in a horizontal direction equal to the velocity at which it left the point of release, since there is no force acting on the ball in a horizontal direction
- undergo constant acceleration downwards in a vertical direction equal to the acceleration due to gravity.

When launched upwards, an object in flight will have a maximum vertical velocity at the point of release. On returning to the same height, the vertical velocity will be the same. It will have a zero vertical velocity at the point where it reaches its maximum height after which it will begin to fall and the velocity will increase.

In the event that a ball is thrown directly horizontally or downwards, the horizontal velocity will still remain constant once the ball is released. The vertical velocity will increase as the ball is accelerated by gravity.

The resultant velocity of the object at any point will be the vector addition of the horizontal and vertical velocities of the object at that point.

This is shown diagrammatically in Figure 18.4.5. By adding the vertical and horizontal vectors, the resultant velocity of the ball can be determined.

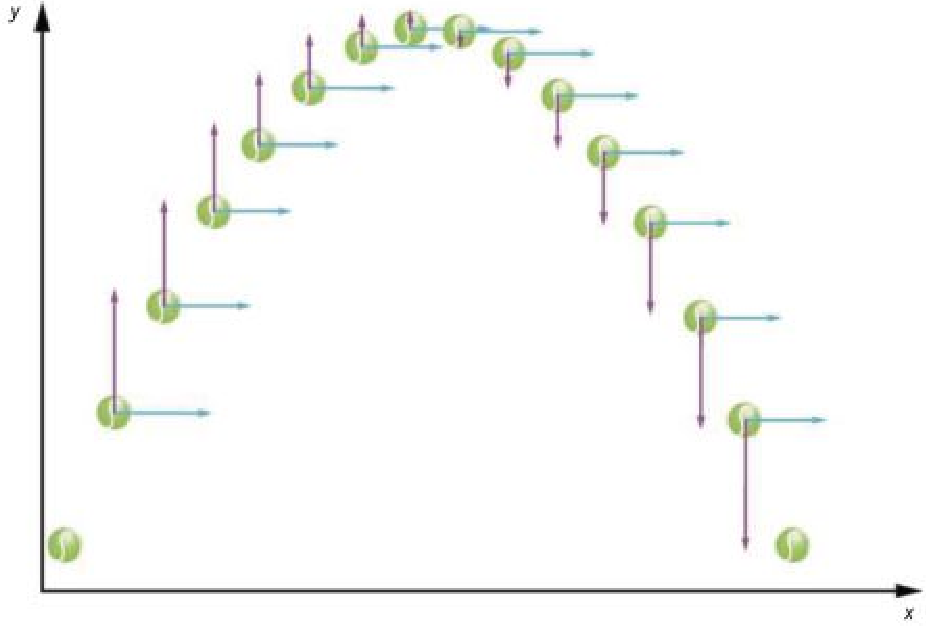


FIGURE 18.4.5 An annotated analysis of the ball from Figure 18.4.2 showing velocity vectors at each point. The horizontal velocity is constant while the vertical velocity vector shows the velocity decreasing while the ball is going up and increasing when the ball is falling.

APPLYING EQUATIONS OF MOTION TO THE FLIGHT OF A BALL

In Chapter 9 'Linear motion', the equations of motion were explained and applied to simple straight-line motion in one dimension. These same equations can be applied to the more complex two-dimensional motion of an object in flight. This is because two-dimensional motion, like that of the bouncing ball, is really no more than the combination of the vertical and horizontal components of the motion working together.

The equations of motion can be simplified for this application based on facts about an object's motion in flight when ignoring air resistance and other external forces other than gravity.

Horizontal motion

In a horizontal direction no forces apply and the acceleration is zero ($a = 0$). Thus:

$v = u + at$ becomes $v = u$

That is, the horizontal velocity remains constant.

Similarly, the other equations simplify to:

$s = ut = vt$

where s is the horizontal distance travelled during flight in metres, often called the *range*

t is the time of flight in seconds.

Vertical motion

In the vertical direction the constant acceleration included in the straight line motion equations is that due to gravity, denoted g and equal to 9.8 m s^{-2}.

The equations of motion can be used in their standard forms, substituting in the constant value of gravity for acceleration.

$$v = u + 9.8t$$
$$v^2 = u^2 + 19.6s$$
$$s = \frac{1}{2}(u + v)t$$
$$s = ut + 4.9t^2$$

Note that the acceleration due to gravity will always be acting downwards. Make sure to label the direction conventions you use and then use these consistently throughout any working.

As noted previously, the resultant velocity of the object at any point will be the vector addition of the horizontal and vertical components.

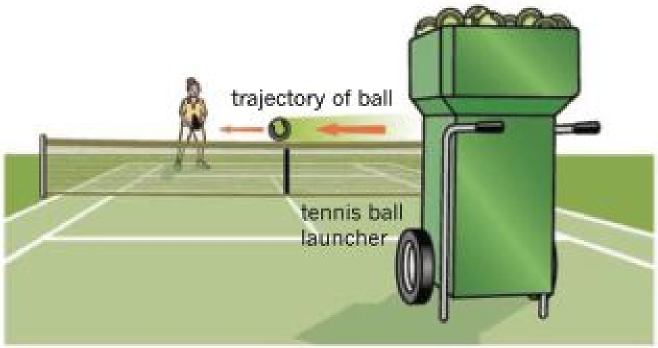
The following worked examples demonstrate how to approach specific types of problems associated with the flight of balls.

Worked example 18.4.1

BALLS HIT OR THROWN HORIZONTALLY

A tennis ball of mass 59 g is being projected horizontally from a practice machine. The initial speed is 30 m s^{-1} and the ball is released 2.0 m from the ground.

a How long, in seconds, will the ball be in the air before hitting the surface of the court?



Thinking	Working
Identify the known information for both the horizontal and vertical components of the ball's motion. As the ball is falling down, it is easiest to take 'down' as positive in this question.	$u_h = 30$ m s^{-1} $u_v = 0$ m s^{-1} $a_h = 0$ $a_v = +9.8$ m s^{-2} $s_h = ?$ $s_v = 2.0$ m $t = ?$
The time the ball takes to fall to the surface of the court will depend on the vertical motion only. From the information supplied, identify the most suitable equation of motion.	In the vertical direction $s = 2.0$ m $u_v = 0$ m s^{-1} $t = ?$ Use the equation: $s = ut + 4.9t^2$
Substitute the known information and rearrange to find t.	$s = ut + 4.9t^2$ $2.0 = 0t + 4.9t^2$ $t^2 = \frac{2.0}{4.9}$ $t = \sqrt{\frac{2.0}{4.9}}$ $= 0.639$ $= 0.64$ s

b Calculate the horizontal distance the ball has travelled while it falls.

Thinking	**Working**
The time the ball takes to fall will be the same time that the ball is travelling horizontally. Use the time and identify known information in the horizontal direction to calculate the distance.	$s_h = ?$ $u_h = 30 \text{ m s}^{-1}$ $a_h = 0$ $t = 0.64 \text{ s}$ The only equation required for calculations in the horizontal direction is $s = ut$ since a is zero and the speed is constant.
Substitute the known information and solve to find s.	$s = ut$ $s = 30 \times 0.64$ $= 19.2 \text{ m}$ $= 19 \text{ m}$

c Calculate the resultant velocity of the ball on landing.

Thinking	**Working**
This calculation requires both the horizontal and vertical components of the ball's velocity to be known. The final horizontal component of the velocity will be the same as the initial, the vertical component will change due to the acceleration due to gravity.	Horizontal: $v = u = 30 \text{ m s}^{-1}$ since a is zero Vertical: $v = ?$ $u = 0$ $s = 2.0 \text{ m}$ $t = 0.64 \text{ s}$ a is positive
Identify a suitable equation based on what is known, substitute and solve for the vertical component of the final velocity. In this case there are two options. Try to use supplied data when possible.	Two options: $v = u + 9.8t$ $v^2 = u^2 + 19.6s$ To make use of supplied data, the best option to use is: $v^2 = u^2 + 19.6s$
Substitute the known information and solve to find v.	$v^2 = u^2 + 19.6s$ $= 0 + 19.6 \times 2.0$ $v = \sqrt{39.2}$ $= 6.26 \text{ m s}^{-1}$
Solve for the final velocity using vector addition of the two components.	
Use Pythagoras's theorem to determine the magnitude of the final resultant velocity.	$v_f^2 = v_h^2 + v_v^2$ $v_f^2 = 30^2 + 6.26^2$ $v_f^2 = 900 + 39.2$ $v_f = \sqrt{939.2}$ $= 30.65$ $= 30.7 \text{ m s}^{-1}$
Use trigonometry to find the angle. Since the adjacent and opposite are known, tan would be the appropriate function.	$\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{v_v}{v_h}$ $\tan\theta = \frac{6.26}{30} = 0.21$, and $\theta = 12°$
State the final velocity giving both the magnitude and the angle relative to the horizontal.	30.7 m s^{-1} at 12° down from the horizontal.

Worked example: Try yourself 18.4.1

BALLS HIT OR THROWN HORIZONTALLY

A tennis ball of mass 59 g is being projected horizontally from a practice machine. The initial speed is 20 m s^{-1} and the ball is released 1.5 m from the ground.

a How long, in seconds, will the ball be in the air before hitting the surface of the court?

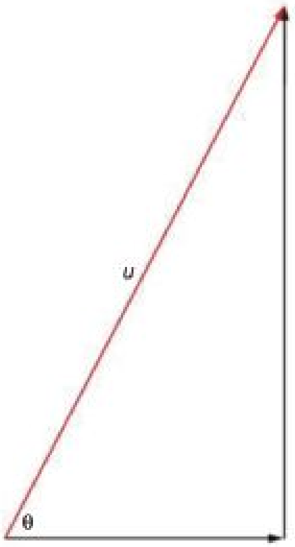
b Calculate the distance the ball has travelled while it falls.

c Calculate the resultant velocity of the ball on landing.

Worked example 18.4.2

BALLS HIT OR THROWN AT AN ANGLE

A netball is thrown upwards with an initial velocity of 15 m s^{-1} at an angle of 60° to the horizontal.

a How long will the ball take to reach its maximum height, assuming no air resistance?	
Thinking	**Working**
The initial values are supplied as a velocity with magnitude and direction. Find the corresponding initial vertical and horizontal components by drawing a triangle and using trigonometry.	 $u_v = u \times \sin\theta = 15 \times \sin 60°$ $= 13$ m s^{-1} $u_h = u \times \cos\theta = 15 \times \cos 60°$ $= 7.5$ m s^{-1}
Time taken to reach the maximum height will be in the vertical direction. Use the vertical component; *v* at the top of the flight will be zero. As the ball is rising, it is easiest to take 'down' as negative and 'up' as positive in this question.	Use the equation: $v = u + 9.8\,t$ Since then ball is rising, take acceleration as negative and the equation becomes: $v = u - 9.8\,t$ $0 = 13 - 9.8\,t$ $t = \frac{-13}{-9.8} = 1.327$ $= 1.3$ s

b Assuming the ball was released at an initial height of 2.0 m, what is the maximum height the ball will reach?

Thinking	Working
Maximum height equals the starting height plus the additional height the ball achieves after release. Use the unrounded value calculated for t in calculating the height to avoid rounding errors.	In the vertical direction, use the equation: $s = ut + 4.9t^2$ Since then ball is rising, take acceleration as negative. The equation becomes: $s = ut - 4.9t^2$ $= 13 \times 1.327 - 4.9 \times 1.327^2$ $= 17.251 - 8.629$ $= 8.62$ m up from the starting point
You may wish to check your working with a second equation applicable for this situation. Recall that the vertical component of velocity will be zero at maximum height.	$s = \frac{1}{2}(u + v)t$ $= \frac{1}{2}(13 + 0) \times 1.327$ $= 8.62$ m
Calculate the final height from the original height plus the additional height calculated after release.	$s = 2.0 + 8.62 = 10.62 = 10.6$ m (up) The ball has been thrown a little too hard for your average netball game!

Worked example: Try yourself 18.4.2

BALLS HIT OR THROWN AT AN ANGLE

A football is kicked upwards with an initial velocity of 20 m s^{-1} at an angle of 45° to the horizontal.

a How long will the ball take to reach its maximum height, assuming no air resistance?

b Assuming the ball was kicked from an initial height of 1.0 m, what is the maximum height the ball will reach?

18.4 Review

SUMMARY

- Ignoring air resistance, a ball or object in flight will:
  - continue at a constant velocity in a horizontal direction equal to the velocity at which it left the point of release. This is because there is no force acting on the object in the horizontal direction.
  - undergo constant acceleration downwards in a vertical direction equal to the acceleration due to gravity.
- When thrown upwards, an object will have a maximum velocity equal to that at the point of release. On returning to the same height the vertical velocity will be the same. It will have a zero vertical velocity at the point where it reaches its maximum height.
- In the event that a ball is thrown directly horizontally or downwards, then the horizontal velocity will remain constant once the ball is released. The vertical velocity will increase as the ball is accelerated by gravity.
- The resultant velocity of the object at any point will be the vector addition of the horizontal and vertical velocities of the object at that point.
- In the vertical direction the constant acceleration used in the straight line motion equations is that due to gravity, denoted *g*, and is equal to 9.8 m s^{-2}.
- In the horizontal direction, the straight line equations of motion simplify to $s = ut = vt$
- In the vertical direction the equations of motion can be used in their standard forms, substituting in the constant value of gravity, 9.8 m s^{-2}, for acceleration.
  - $v = u + 9.8t$
  - $v^2 = u^2 + 19.6s$
  - $s = \frac{1}{2}(u + v)t$
  - $s = ut + 4.9t^2$
- The acceleration due to gravity will always be acting downwards. If a ball or object is rising it is easiest to take the acceleration and other downwards measurements as negative, if the object is falling then it is easiest to take 'down' as positive. It is important to indicate the convention used and to use it consistently throughout the question.

KEY QUESTIONS

1 Ignoring air resistance, describe the horizontal velocity of a basketball as it is thrown towards the hoop in a shot for goal.

2 Ignoring air resistance, when will the vertical velocity of a ball in flight be at a minimum?

3 A golfer hits a ball from a tee. The ball lofts into the air and returns to the ground at the same horizontal height some distance down the fairway. Ignoring air resistance, which of the following best describes the comparison between the time that the ball takes to rise to its maximum height during its trajectory and the time it takes to fall back to the ground?

A The time to rise is more than the time to fall.
B The time to rise is less than the time to fall.
C The time to rise is the same as the time to fall.
D A comparison cannot be drawn from the information given.

The following information relates to questions 4 to 6.

A golfer hits a ball from the tee with an initial velocity such that the vertical component is 20 m s^{-1} and horizontal component is 30 m s^{-1}. Assume no air resistance and that the launch and landing heights are the same.

4 What will be the vertical and horizontal components of the ball's velocity at the point just before landing?

5 What will the value of the vertical component of the velocity be 1.5 seconds after leaving the tee?

6 **a** How long will the golf ball be in the air before landing?
b What horizontal distance does the ball travel?

7 A netballer stands almost directly under the ring when shooting for a goal. She holds the ball at a height of 1.80 metres and then shoots for goal. In order for the ball to clear the ring and drop in, it must reach a height of at least 3.30 metres. What is the minimum speed, ignoring air resistance, that the ball must leave her hands to score a goal? Use $g = 9.8$ m s^{-2} and give your answer to three significant figures.

8 An AFL player kicks a high-lofting torpedo punt towards the goal from the 50 metre line. It leaves his foot at an angle of 45° from the horizontal and with a speed of 23 m s^{-1}. Assuming no air resistance, calculate the total horizontal distance the ball travels. Give your answer to two significant figures.

18.5 Air resistance

In order to simplify the calculations associated with projectile motion, scientists often assume that air resistance is negligible. While this is a useful approximation, there are many situations where the effect of air resistance is significant.

In ball sports, competitors will often use the effects of air resistance to control how the ball moves through the air. For example, a cricketer, footballer or tennis player will often spin the ball to cause it to dip or curve as it moves through the air. Even a small amount of change from the ball's expected parabolic path can sometimes create an advantage (see Figure 18.5.1).

FIGURE 18.5.1 Table tennis players often impart a significant amount of spin on the ball to cause it to dip or curve.

THE EFFECT OF AIR RESISTANCE

To simplify physics calculations, the effects of air resistance are often ignored. However, in reality, any ball that moves through the air experiences a retarding force due to fluid friction. This force is called **drag**.

When a person throws a ball horizontally, the drag force from the air acts in the opposite direction to the ball's velocity.

Without drag, the ball would follow a perfectly parabolic path; however, since the additional drag force acts upwards and backwards, this reduces the range (i.e. horizontal distance) of the ball and causes it to 'hang' in the air a little bit longer than it would without air resistance. The shape of the new path followed by the ball is known as a 'ballistic curve' and is shown in Figure 18.5.2.

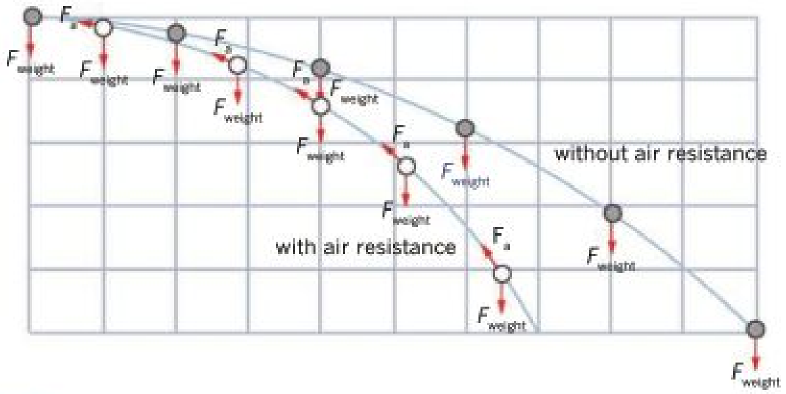


FIGURE 18.5.2 The ballistic curve of the ball with air resistance highlights the reduced range due to the retarding force.

Calculating drag

It is difficult to mathematically determine the path of a ball affected by air resistance because, unlike gravity, drag is not a constant force. The amount of drag varies with the speed of the ball: the faster the ball, the more drag it experiences. In fact, drag varies with the square of the ball's speed, i.e. $F_D \propto v^2$. Other factors that affect the drag experienced by a ball include the density of the air and the area and shape of the ball.

Mathematically, the drag force, F_D (in N), can be calculated using the equation:

$$F_D = \frac{1}{2}C_D\rho v^2 A$$

where v is the velocity of the ball (in m s^{-1})

ρ is the density of the air (in kg m^{-3})

A is the cross-sectional area of the ball (in m^2).

C_D is a dimensionless constant known as the *drag coefficient* which depends on the shape of the ball. This is usually determined by conducting experiments in a wind tunnel. For a smooth ball C_D is approximately 0.5, but most sports balls are not perfectly smooth due to stitching or ripples on their surfaces. This roughness actually reduces drag by creating a layer of air that sits around the ball and allows it to move more smoothly through the air. Typically, baseballs and cricket balls have drag coefficient of around 0.35, whereas a golf ball has a coefficient of less than 0.2.

Worked example 18.5.1

CALCULATING DRAG

Calculate the drag force acting on a cricket ball (diameter 72.6 mm) travelling at 40 m s^{-1} if its drag coefficient is 0.33 and the air density is 1.23 kg m^{-3}.

Thinking	Working
Calculate the cross-sectional area of the ball.	$D = 72.6$ mm $= 0.0726$ m $\therefore r = \frac{0.0726}{2} = 0.0363$ m $A = \pi r^2 = \pi \times 0.0363^2$ $= 0.00414$ m^2
Recall the formula for drag.	$F_D = \frac{1}{2}C_D\rho v^2 A$
Calculate the drag force.	$F_D = \frac{1}{2} \times 0.33 \times 1.23 \times 40^2 \times 0.00414$ $= 1.3$ N

Worked example: Try yourself 18.5.1

CALCULATING DRAG

Calculate the drag force acting on a baseball travelling at 47 m s^{-1} if it has a diameter of 72.1 mm and a drag coefficient of 0.36. Assume an air density of 1.25 kg m^{-3}. Give your answer correct to two significant figures.

Terminal velocity

Consider a ball that has been dropped from a great height. As the ball accelerates towards the ground, its velocity increases and the drag force acting on it increases dramatically. Eventually, the drag force will be equal to the weight of the ball. At this point, the ball will stop accelerating and will travel at a constant speed. This maximum constant falling speed is known as the ball's **terminal velocity**.

At terminal velocity: $F_D = F_g$

Therefore, for a ball:

$$\frac{1}{2}C_D\rho v^2 A = mg$$

$$v_t^2 = \frac{2mg}{C_D\rho A}$$

$$v_t = \sqrt{\frac{2mg}{C_D\rho A}}$$

where v_t is terminal velocity (in m s^{-1})

m is the mass of the falling object (in kg)

g is acceleration due to gravity (9.8 m s^{-2} down).

Worked example 18.5.2

CALCULATING TERMINAL VELOCITY

A cricket ball has a mass of 161 g and a diameter 72.6 mm. Calculate its terminal velocity assuming a drag coefficient of 0.33, an air density of 1.15 kg m^{-3} and gravitational acceleration of 9.8 m s^{-2}.

Thinking	Working
Calculate the cross-sectional area of the ball.	$D = 72.6 \text{ mm} = 0.0726 \text{ m}$ $\therefore r = \frac{0.0726}{2} = 0.0363 \text{ m}$ $A = \pi r^2 = \pi \times 0.0363^2$ $= 0.00414 \text{ m}^2$
Recall the formula for terminal velocity.	$v_t = \sqrt{\frac{2mg}{C_D\rho A}}$
Calculate the terminal velocity.	$v_t = \sqrt{\frac{2 \times 0.161 \times 9.8}{0.33 \times 1.15 \times 0.00414}}$ $v_t = 45 \text{ m s}^{-1}$

Worked example: Try yourself 18.5.2

CALCULATING TERMINAL VELOCITY

Calculate the terminal velocity of a tennis ball (mass 58.9 g and diameter 67 mm) assuming gravitational acceleration is 9.8 m s^{-2}, a drag coefficient of 0.51 and air density of 1.26 kg m^{-3}. Give your answer correct to two significant figures.

WHY SOME BALLS ‘SWING’ IN THE AIR

It is possible to get a ball to move unpredictably as it flies through the air. This is what spin bowlers in cricket refer to as drift and pace bowlers call swing. Similar techniques are used in sports such as baseball and soccer.

Swing through drag

Perhaps the simplest method to create swing is to shine one side of a cricket ball. As a cricket ball is used it becomes worn. During a cricket match, players on the fielding side will choose one side of the ball that they will try to keep as shiny as possible. They will often do this by applying their sweat or spit to the ball and then polishing it on their trousers (hence the red marks often seen on the trousers of fast bowlers). When the ball is bowled, the polished side experiences less fluid friction or drag with the air than the rough side and the ball swings towards the rough side (as seen in Figures 18.5.3 and 18.5.4 on page 634). Since the amount of drag on an object is proportional to the density of the fluid through which the object moves, this effect will be most pronounced when the density of air is highest. This is why bowlers get much more swing in humid weather than in dry weather. High humidity means there is a lot of water vapour in the air. This makes the air denser.

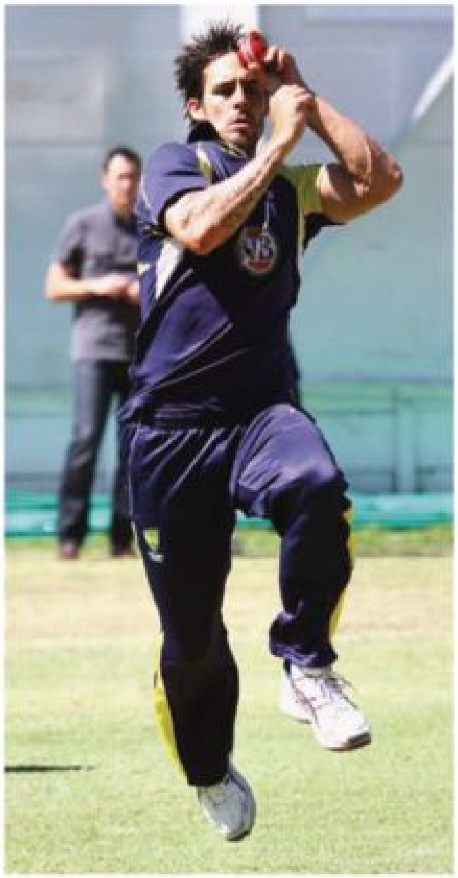


FIGURE 18.5.4 Seam bowler Mitchell Johnson holds the ball with the seam upright so that it will swing.

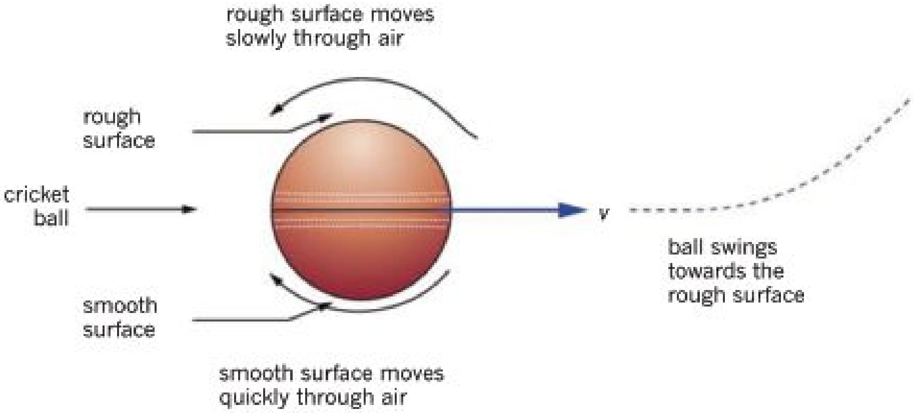


FIGURE 18.5.3 Bowlers often shine up the ball on one side so that it will curve to the rough side when it is bowled.

Magnus force

Another way to get movement through the air is using Bernoulli's principle. According to Bernoulli's principle, if air flows more quickly over one surface of the ball than another, it creates an area of low pressure on the side of the ball where the air is moving faster. The difference in pressure will cause the ball to move towards the low-pressure region. This will make the ball curve in mid-air.

Sports people use a variety of different techniques to make air travel at different speeds over each surface of a ball.

Spin bowlers can take advantage of Bernoulli's principle by using what is known as the Magnus force (Figure 18.5.5). When a ball moves through the air, a small layer of air called the boundary layer moves along with it. If the ball is spinning, the boundary layer moves with the surface of the ball. The velocity of this boundary layer combines with the velocity of the air rushing past the ball as it moves forwards. This means that on one side of the ball, the air is moving faster than on the other side. Bernoulli's principle applies and the ball moves towards the region of lower pressure (i.e. where the air is moving faster).

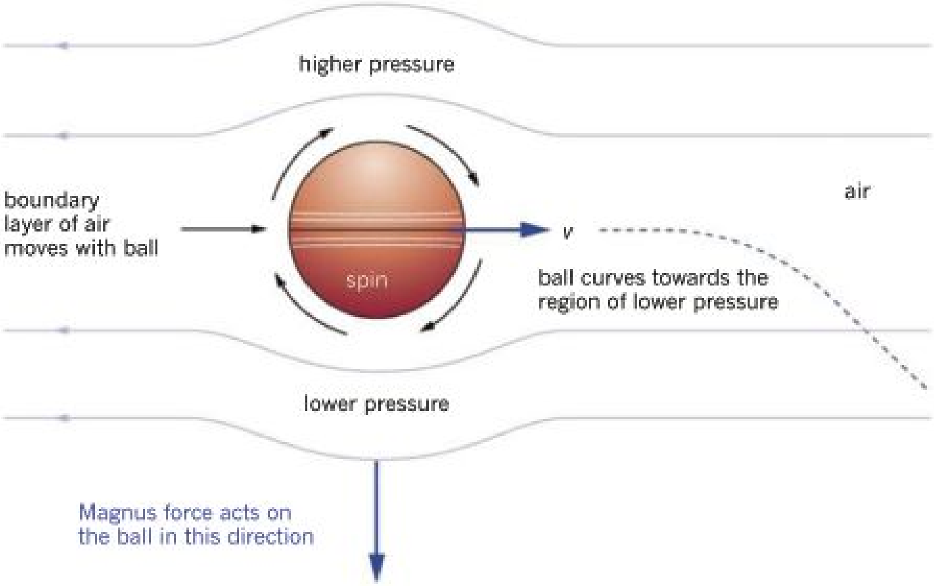


FIGURE 18.5.5 When a ball moves through the air, the boundary layer of air moves with it, creating the Magnus effect.

18.5 Review

SUMMARY

- Air resistance causes balls to follow a trajectory known as a ballistics curve rather than a parabolic path.
- Drag force can be calculated using the equation:

$$F_D = \frac{1}{2}C_D\rho v^2 A$$

- When the drag force on a ball is equal to the weight of the ball, the ball will stop accelerating and will travel at a constant speed known as its terminal velocity.
- Terminal velocity can be calculated using the equation:

$$v_t = \sqrt{\frac{2mg}{C_D\rho A}}$$

- A spinning ball experiences a Magnus force because the air pressure one side of the ball is different to the air pressure on the other side of the ball.

KEY QUESTIONS

1 Complete the following sentence:
Often calculations of the trajectory of a ball ignore air resistance. When compared to the ideal parabolic path, a ball that is affected by air resistance:

A travels faster
B reaches a higher maximum height
C travels a longer horizontal distance
D travels a shorter horizontal distance

2 Which of the following statements is true?

A Two balls of the same diameter, one made of lead and the other made of polystyrene, are dropped from a tower. If you disregard drag, the velocity of the lead ball will be greater than the velocity of the polystyrene ball.
B The force of the drag F_D will keep increasing until it is equal to the force of gravity of the falling body.
C If you take drag into consideration the velocity of the falling ball will keep increasing linearly.
D The drag affecting a falling ball is constantly equal to the force of gravity.

3 The drag experienced by an object will depend on the medium through which it is travelling. Put the following mediums in order from the one where the greatest drag is experienced to the one where least drag is experienced.

air
vacuum
water
glycerine

4 Determine the missing values from the table below. Give your answers correct to two significant figures.

Ball type	Diameter (mm)	Radius (m)	Cross-sectional area (m^2)
netball	226 mm		
cricket		0.036	
tennis			0.0036

5 Calculate the drag force acting on a softball travelling at 30 m s^{-1} if it has a diameter of 76 mm and a drag coefficient of 0.35. Assume an air density of 1.25 kg m^{-3}. Give your answer correct to two significant figures.

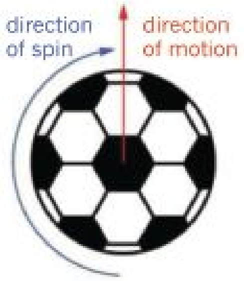
6 A cricket ball has a mass of 160 g and a diameter of 72.6 mm. Calculate the change in its terminal velocity if its drag coefficient changes from 0.33 to 0.4 during a match due to scuffing. Assume the air density is 1.25 kg m^{-3} and gravitational acceleration is 9.8 m s^{-2}. Give your answer correct to two significant figures.

7 Calculate the terminal velocity in air (density 1.2 kg m^{-3}) of a smooth rubber ball that has a mass of 200 g, a diameter of 100 mm and a drag coefficient of 0.45. Assume gravitational acceleration is 9.8 m s^{-2}. State your answer correct to two significant figures.

8 Calculate the drag coefficient in air of a softball. The ball has a mass of 195 g, a diameter of 98 mm and a terminal velocity of 39 m s^{-1}. Assume gravitational acceleration is 9.8 m s^{-2} and air density is 1.25 kg m^{-3}. Give your answer correct to two significant figures.

9 A cricket ball (m = 161 g, C_D = 0.33) and tennis ball (m = 59 g, C_D = 0.51) are approximately the same size and shape. Explain why the cricket ball's terminal velocity (40 m s^{-1}) is much greater than that of the tennis ball (30 m s^{-1}).

10 The diagram below shows a football that has been kicked in the direction indicated by the red arrow. The blue arrow indicates the way the ball is spinning. Due to this spin, the ball will experience a Magnus force. Copy the diagram and draw an arrow showing the direction of the Magnus force.



Chapter review

18

KEY TERMS

angular speed
coefficient of restitution
drag
elastic
elastic collision
inelastic collision
kinetic friction
pendulum
point mass
radian
rotational kinetic energy
rotational motion
simple harmonic motion
static friction
terminal velocity
translational kinetic energy
translational motion

1 A football has a coefficient of restitution of 0.75. What height will it bounce back to if it is dropped from a height of 1.00 m?

A 0.56 m
B 0.75 m
C 1.00 m
D 1.78 m

2 A netball rebounds at 25 m s^{-1} after it hits the ground. If it has a coefficient of restitution of 0.78, how much speed did it lose in the bounce?

3 A 200 g snooker ball with initial velocity 9.0 m s^{-1} to the right collides with a stationary snooker ball of mass 100 g. After the collision, both balls are moving to the right. The 200 g ball has a speed of 3.0 m s^{-1}, while the 100 g ball has a speed of 12.0 m s^{-1}.

a Calculate the total kinetic energy of the system before the collision.
b Calculate the total kinetic energy of the system after the collision.
c Determine if the collision is elastic, inelastic or impossible.

4 Two identical bowling balls, each of mass 4.0 kg, move towards each other across a frictionless horizontal surface with equal speeds of 3.0 m s^{-1}. During the collision, 20 J of the initial kinetic energy is transformed into heat and sound. After the collision the balls move in opposite directions away from each other.

a Is momentum conserved in this collision?
b Is this an elastic or inelastic collision?
c Calculate the speed of each ball after the collision.

5 A ball of mass m slides along the flat ground with a coefficient of friction of μ for a distance s. The initial speed of the ball is u and the acceleration due to gravity is g. Find an expression for the deceleration of the ball in terms of μ and g.

6 A ball with a mass of 425 g skids along the ground experiencing a frictional force of 2.0 N. Calculate the coefficient of friction between the ball and the ground.

7 If a netball with a mass of 420 g experiences a coefficient of friction of 0.52 as it skids along the court, calculate the force of friction acting on the ball.

8 A bocce ball has a diameter of 112 mm. When it is bowled along the ground it has an initial speed of 10.9 m s^{-1} before coming to rest in 4.11 s. Calculate the following, giving your answers correct to three significant figures:

a the total distance travelled
b the average linear and angular speed of the ball.

9 An astronaut finds that a 60 cm pendulum takes 25.2 seconds to complete 10 swings on Mars.

a Use this information to determine the value of g on Mars. Give your answer correct to three significant figures.
b How much lower, as a percentage, is this compared to Earth where $g = 9.8$ m s^{-2}. Give your answer correct to two significant figures.

10 Complete the paragraph by choosing the correct response from the choices provided in brackets.

When a golfer swings the club from above her head to strike the ball [**elastic/gravitational potential/kinetic**] energy will be converted into [**elastic/gravitational potential/kinetic**] energy. This system can be best modelled as a [**double pendulum/falling object/simple pendulum**].

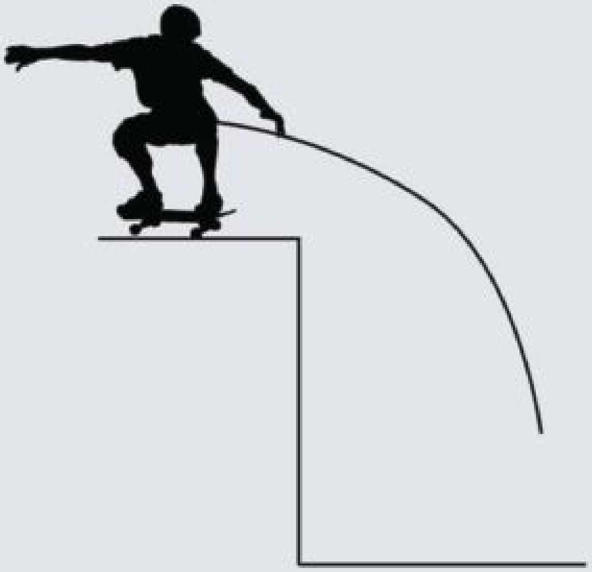
11 A batsman lifts his 1.3 kg bat off the ground to get ready to face a ball. The centre of mass of the bat is raised by 1.4 m. Identify the amount of gravitational potential energy gained by the bat. Give your answer correct to three significant figures.

12 The batsman from question 11 now swings at an oncoming ball and misses. It is found that at the bottom of the bat's swing, all of the gravitational potential energy is converted into kinetic energy. Find the maximum speed of the bat when the bat's centre of mass is at 0 m, i.e. at the bottom of the swing. Assume no extra force is given to the swing.

13 A baseball spectator in the terrace throws a ball back to a player on the field. The ball leaves the spectators hand moving horizontally at 40 m s^{-1} at a vertical height of 17.5 m above the ground. Ignore air resistance and give you answer correct to two significant figures.

- **a** Calculate the time it takes for the baseball to strike the ground.
- **b** How far does the baseball travel horizontally before it hits the ground?
- **c** What is the speed of the baseball as it reaches the ground?

14 A skate boarder takes off travelling horizontally at 5 m s^{-1} from a bank. She follows a parabolic path before landing safely onto the level below. The flight takes 0.8 seconds.

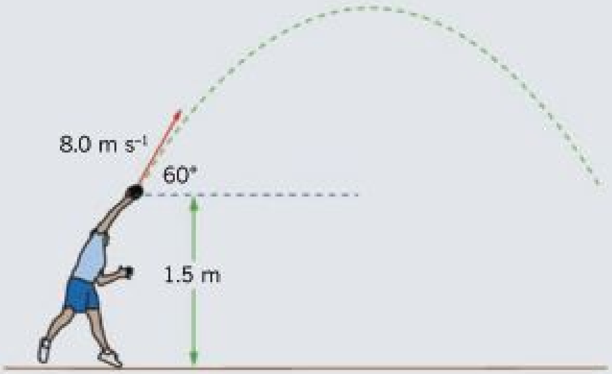


- **a** What is the initial vertical velocity of the skateboarder?
- **b** What is the vertical component of the skater's velocity just before she hits the ground?
- **c** How far does the skater fall vertically?
- **d** What is the speed of the skater just before she hits the ground?
- **e** At what angle relative to the horizontal is the skate boarder travelling as she reaches the ground?

15 An AFL player kicks a ball at 20 m s^{-1} and at 65° to the horizontal over an opposition player. It is marked at the same height as it was kicked by another team member further up the ground. Use g = 9.8 m s^{-2} and ignore air resistance. Give your answer for each of the following questions correct to three significant figures.

- **a** How long will the ball be in the air before it is marked?
- **b** How far does the ball travel horizontally before it is marked?

16 In a shot-put event a 2.0 kg shot is launched from a height of 1.5 m, with an initial velocity of 8.0 m s^{-1} at an angle of 60° to the horizontal. Give your answer to each of the following questions correct to two significant figures.



- **a** What is the initial horizontal speed of the shot?
- **b** What is the initial vertical speed of the shot?
- **c** How long does it take the shot-put to reach its maximum height?
- **d** What is the maximum height from the ground that is reached by the shot?
- **e** What is the speed of the shot when it reaches its maximum height?

17 Calculate the drag force acting on a tennis ball travelling at 55 m s^{-1} if it has a diameter of 68 mm and a drag coefficient of 0.56. Assume an air density of 1.25 kg m^{-3}. Give your answer correct to two significant figures.

18 Air temperature is known to affect the density of air. Normally air density can be assumed to equal 1.25 kg m^{-3}. At 35°C the air density is 1.15 kg m^{-3}. Find the difference in the terminal velocity that this change in air density would have on a tennis ball (mass 57.2 g and diameter 66.5 mm). Assume a drag coefficient of 0.49 and gravitational acceleration of 9.8 m s^{-2}. Give your answer correct to two significant figures.

19 Explain why air resistance reduces the range of a golf shot.

20 The Magnus effect, named after the German physicist Heinrich Gustav Magnus (who first explored the phenomenon experimentally) explains how projectiles can curve when moving through a fluid like air. Whenever a ball is spinning through the air, the Magnus effect has consequences for

- air pressure
- direction of ball movement and
- distance travelled compared to a non-spinning flight path.

Using these three points explain what effect topspin has on a tennis ball.

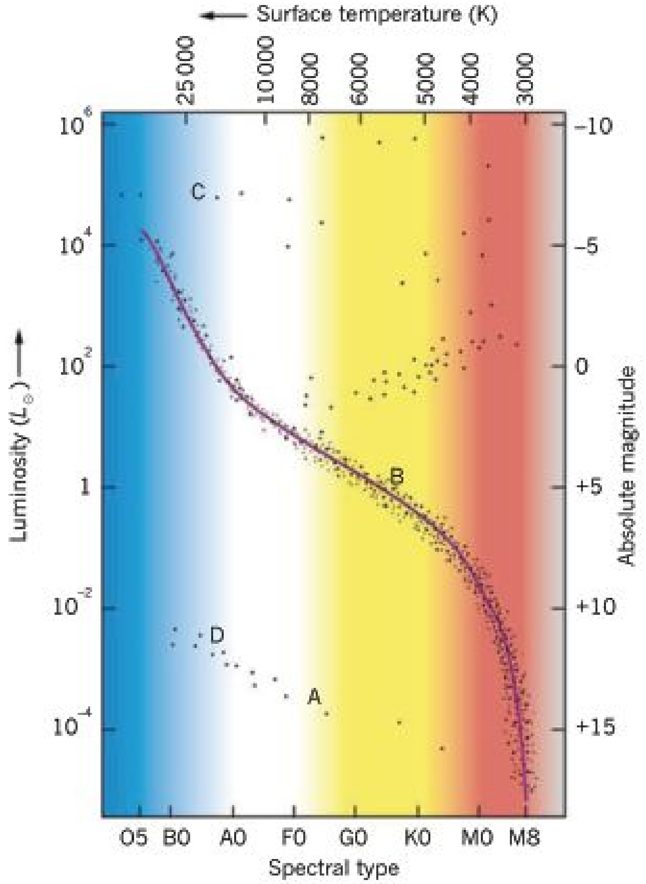
REVIEW QUESTIONS

Options

How can thermal effects be explained?

Stars

The following information relates to questions 1 to 3. Consider the Hertzsprung–Russell diagram shown below.



1 Which letter corresponds to a Sun-like star?
A A
B B
C C
D D

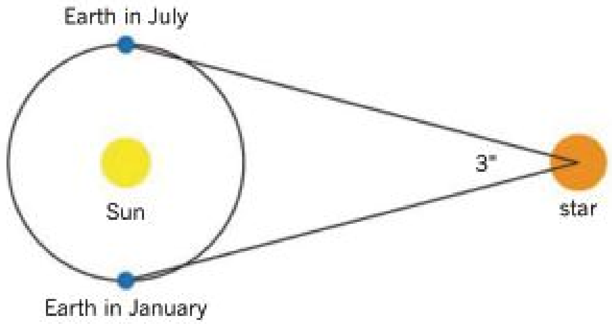
2 Which letter corresponds to a blue supergiant?
A A
B B
C C
D D

3 Which letter corresponds to an old star that once was a Sun-like, main sequence star?
A A
B B
C C
D D

4 Which of the following best outlines the life cycle for a massive star, a star with a mass much greater than that of the Sun?
A main sequence star → planetary nebula → red supergiant → supernova → black hole
B main sequence star → planetary nebula → supernova → red supergiant → black hole
C planetary nebula → main sequence star → supernova → red supergiant → black hole
D planetary nebula → main sequence star → red supergiant → supernova → black hole

5 If star A and star B are the same luminosity, but star A is 4 times farther than star B, how do their apparent brightness compare?
A Star A's apparent brightness is 4 times greater than that of star B.
B Star B's apparent brightness is 4 times greater than that of star A.
C Star A's apparent brightness is 16 times greater than that of star B.
D Star B's apparent brightness is 16 times greater than that of star A.

6 The following diagram depicts the concept of parallax.

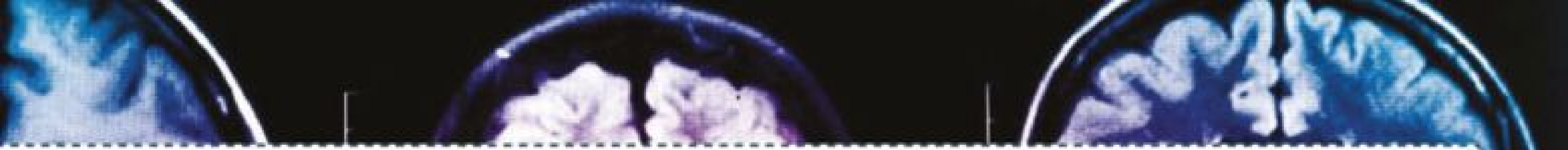


If the full angle subtended by the Earth's January and July positions is 3" (as marked), determine the distance from the Earth to the star in:
a parsecs
b AU

7 The event horizon, the space around a black hole, has a radius described by Schwarzschild's equation:

$$r_s = \frac{2GM}{c^2}$$

a Explain what is meant by the 'event horizon'.
b The mass of our Sun is 2.0×10^{30} kg.
The mass of the star Betelgeuse is not known exactly, but is thought to fall somewhere between 7.7 solar masses and 20 solar masses.
Determine the range of distances (radii) in km that the Schwarzschild radius would be for Betelgeuse.



8 Outline the main ideas presented in this chapter regarding Einstein's idea of 'spacetime'.

9 Define absolute magnitude in terms of a star's brightness. Explain how absolute magnitude and apparent magnitude are related.

10 The luminosity, L, of a star can be determined by the Stefan–Boltzman law:

$L = \sigma \times T^4 \times$ surface area of star

where σ is the Stefan–Boltzman constant, 5.67×10^{-8} W m^{-2} K^{-4}

and T is the surface temperature of the star.

Remember that the surface area of a sphere is $4\pi r^2$.

The surface temperature, T, of the Sun is 5800 K and its radius is 6.96×10^8 m.

a Calculate the luminosity of the Sun.

b Explain what change would be observed in the luminosity of the Sun if its radius was twice its current value.

c Explain what change would be observed from the original luminosity of the Sun if the surface temperature was halved.

Forces in the human body

11 Which of the following statements best describes the situation for an object in translational equilibrium.

A The net force on the object is zero.
B The object experiences no acceleration.
C The object is at constant velocity.
D All of the above are correct.

12 The maximum compressive stress for bone is given as approximately 170 MPa. What information does this value provide about the strength of bone?

A 170 N per m^2 of compressive force is required to 'break' a bone sample.
B 170×10^6 N of compressive force is required to 'break' a bone sample.
C 170×10^6 N per cm^2 of compressive force is required to 'break' a bone sample.
D 170×10^6 N per m^2 of compressive force is required to 'break' a bone sample.

13 Bone sample A is loaded with a force of F newtons. A second sample B, with twice the diameter of the first, is also loaded with a force of F newtons.

What is the ratio of stress on bone sample A to stress on bone sample B?

A 1:2
B 2:1
C 4:1
D 1:4

14 A tendon of original length 12.00 cm experiences a 2.50% strain. What is the new length of the tendon in this situation?

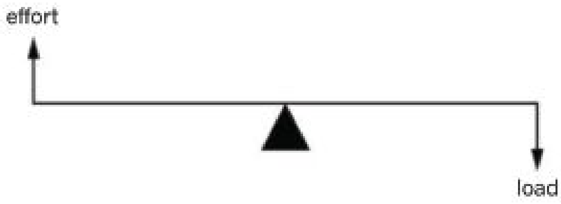
A 11.70 cm
B 11.98 cm
C 12.03 cm
D 12.30 cm

15 As people age, their bones can become brittle. Bones are considered brittle when they:

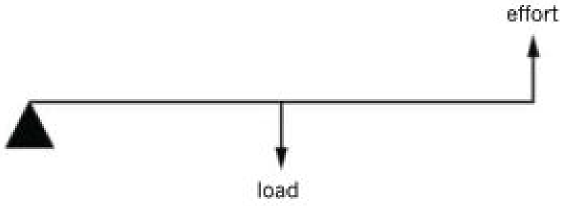
A undergo limited plastic deformation
B undergo limited elastic deformation
C have a large Young's modulus
D absorb a large amount of strain energy before failing

16 Which of the following is an example of a class 2 lever?

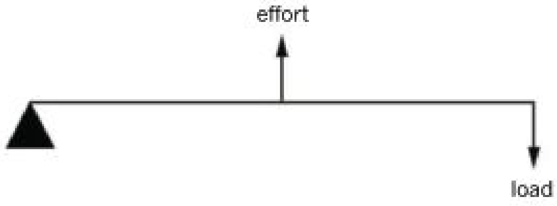
A



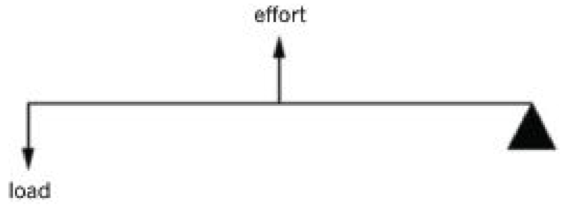
B

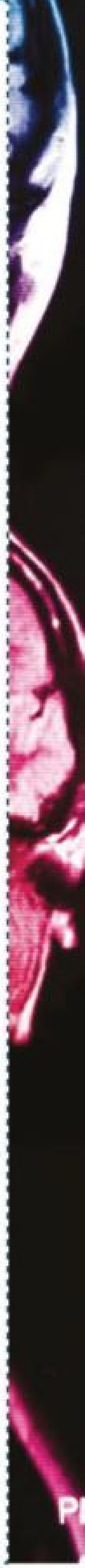


C

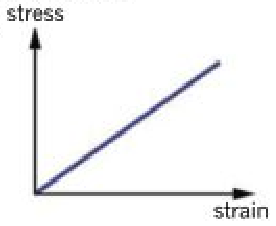


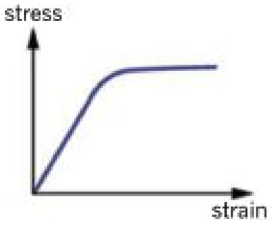
D

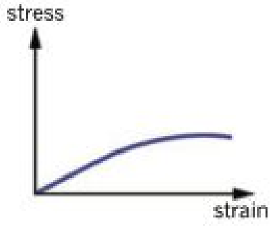


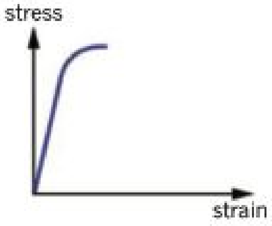


17 Which of the following stress vs strain graphs depicts a material that is stiff, undergoes plastic deformation and is tough?

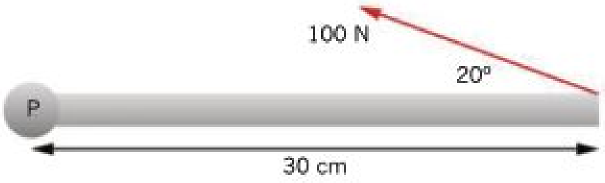
A


B


C


D


18 What is the torque created by the force about point P in the following situation?



A (100 × 30 × sin 20°) N m
B (100 × 30 × cos 20°) N m
C (100 × 0.3 × sin 20°) N m
D (100 × 0.3 × cos 20°) N m

19 What properties of tendons make them suitable for attaching muscles to bones?

A Tendons are very elastic in behaviour and can withstand high tensile loads.
B Tendons are very stiff and can withstand high tensile loads.
C Tendons are brittle and do not undergo much plastic deformation.
D Tendons are very elastic in behaviour and are useful under compressive loads.

20 Which of the following wine glasses would be most stable?

A


B

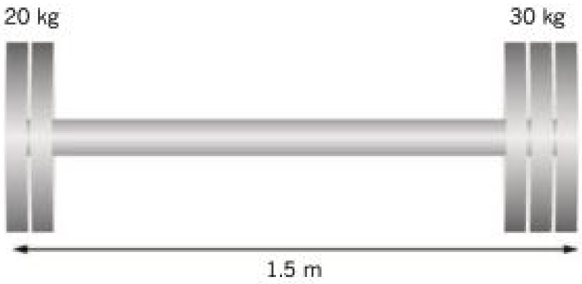

C


D

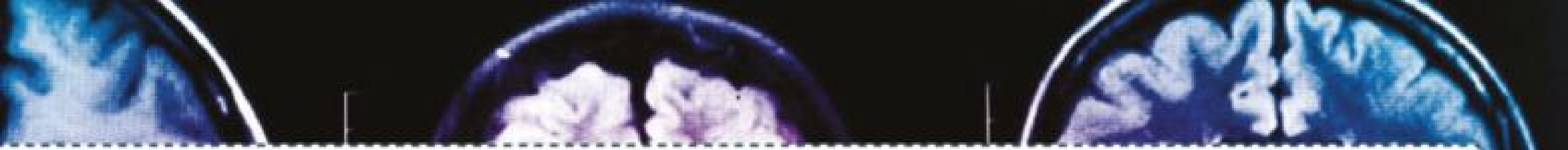

21 Forces applied at two or more points on a body may not cause movement, but cause internal stresses. Describe the three outcomes (types of forces) that two forces may cause.

22 Calculate the compressive stress on a circular bone sample of radius 12 mm when it is exposed to 150 N of compressive force. Give your answer in kPa.

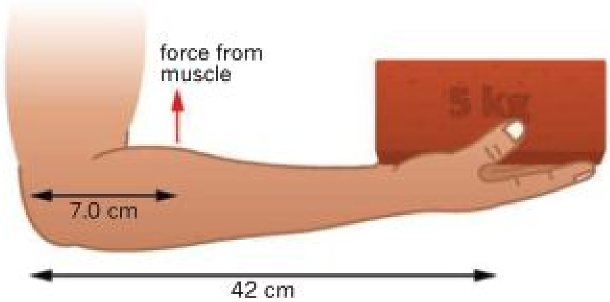
23 An 'uneven' barbell is shown in the diagram below. The mass of the bar itself is 5 kg. Determine the centre of mass of this barbell. Give your answer from the left-hand side. You can assume that the 'weights' are located exactly on each end and that the bar is a uniform material.



24 While kicking a soccer ball, the foot of the soccer player exerts a force of 80 N on the ball as shown in the diagram. Determine the torque exerted by the foot around the knee joint.



25 The forearm can be considered to be a simple lever, as shown in the diagram below. Use $g = 9.8\ \text{N kg}^{-1}$.

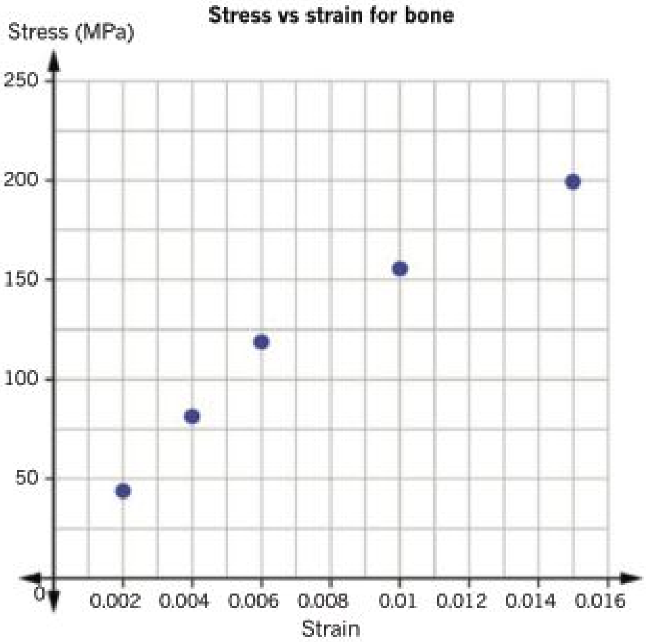


Taking the weight of the forearm (1.0 kg) itself to be acting from the middle of the forearm, determine the vertical force that the muscle must provide to hold a 5.0 kg weight as shown.

26 Explain why a tight rope walker would use a long, flexible pole in order to complete a tight rope walk.

27 A bone sample with a Young's modulus of 10 GPa has a cross-sectional area of $2.5\ \text{cm}^2$ and a length of 20 cm. The sample is subjected to a compressive force of 500 N. What change in length does the bone sample experience?

The following information relates to questions 28 and 29. The following graph shows the relationship between stress and strain for a human bone sample.



28 From the graph, determine Young's modulus for the bone sample.

29 The bone sample has a cross-sectional area of $3.00\ \text{cm}^2$ and a length of 10.0 cm. Calculate the amount of strain energy stored in the bone when it is compressed by 0.60%.

Energy from nuclear power

30 Which of the following components of a fission-powered nuclear reactor is responsible for reducing the energy of emitted neutrons in order to increase the probability of absorption by further nuclei?

A moderator
B heat exchanger
C control rod
D fuel rod

31 Which of the following is a suitable material for the control rods of a fission-powered nuclear reactor?

A water
B boron
C graphite
D uranium

32 The energy output from the Sun involves the fusion of four protons to form one helium nucleus, two antineutrinos, two gamma rays and 24.7 MeV of energy. What is the energy equivalent of 24.7 MeV in joules?

A 3.95×10^{-18} J
B 3.95×10^{-15} J
C 3.95×10^{-12} J
D 3.95×10^{-9} J

33 The amount of nuclear fuel required as a 'critical mass' to allow the fuel to sustain a chain reaction depends on which of the following?

A the mass of the nuclear fuel
B the shape of the nuclear fuel
C the purity of the nuclear fuel
D all of the above

34 The first fission reaction ever to be observed is shown in the equation below:

$^{235}_{92}\text{U} + ^{1}_{0}\text{n} \rightarrow ^{92}_{36}\text{Kr} + ^{141}_{56}\text{Ba} + x^{1}_{0}\text{n} + \text{energy}$

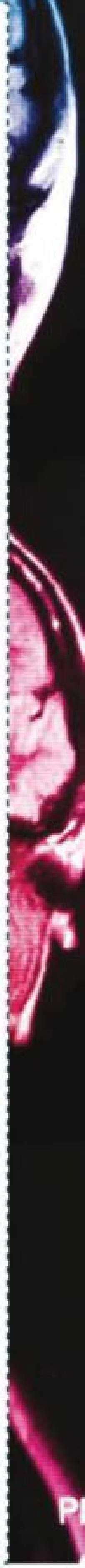
How many neutrons (x) are produced in this fission reaction?

A 1
B 2
C 3
D 4

The following information relates to questions 35 to 37. In one fusion reaction, two deuterium ($^{2}_{1}\text{H}$) nuclei combine to form a helium ($^{4}_{2}\text{He}$) nucleus. The reaction also releases energy.

35 Write a balanced equation for this reaction, including nucleon (mass) and proton numbers.

36 Given that the mass defect for this reaction is 5.92854×10^{-30} kg, calculate the energy (in joules) released in this reaction. Use $c = 3.00 \times 10^{8}\ \text{m s}^{-1}$.



37 Hence calculate the energy (in joules) released when 1.0 kg of deuterium fuses to form helium, $^{3}_{2}He$.
1.0 kg of deuterium contains 3.0×10^{25} deuterium nuclei.

The following information relates to questions 38 and 39. In a nuclear reactor, a U-238 nucleus captures a neutron and becomes U-239. U-239 then decays to Pu-239 through the process of two β^- decays.

38 **a** Write a balanced equation to show the neutron capture by a U-238 nucleus.
b Write two further equations that show the decay of U-238 to Pu-239.
(Note: you will need to refer to a periodic table to complete this question.)

39 What type of reactor makes use of uranium-238 and fast neutrons?

Nuclear medicine

40 Which of the following lists only includes non-ionising radiation?
- **A** infrared radiation, alpha particles, beta particles, X-rays
- **B** alpha particles, beta particles, X-rays, gamma rays
- **C** microwaves, infrared radiation, visible light
- **D** infrared radiation, alpha particles, beta particles, gamma rays

41 What is the difference between soft X-rays and hard X-rays?
- **A** Soft X-rays have a shorter wavelength than hard X-rays.
- **B** Soft X-rays have a lower frequency than hard X-rays.
- **C** Soft X-rays are more energetic than X-rays.
- **D** Soft X-rays are non-ionising while hard X-rays are ionising.

42 Which of the following statements about the effects of radiation is correct?
- **A** 0.1 Sv of beta radiation is biologically as damaging as 2 Sv of alpha radiation.
- **B** 1 Sv of beta radiation is biologically as damaging as 1 Sv of alpha radiation.
- **C** 1 Gy of beta radiation is biologically as damaging as 1 Gy of alpha radiation.
- **D** 0.1 Gy of beta radiation is biologically as damaging as 2 Gy of alpha radiation.

43 Which of the following lists include only techniques that are used for diagnostic purposes?
- **A** PET scans, gamma knife, brachytherapy, SPECT scans
- **B** MRI, chemotherapy, CT scans, brachytherapy
- **C** PET scans, MRI, X-rays, brachytherapy,
- **D** CT scans, PET scans, MRI, X-rays

44 Which of the following lists radiation in increasing order of ionising ability?
- **A** gamma rays, X-rays, beta particles, alpha particles
- **B** X-rays, gamma rays, beta particles, alpha particles
- **C** X-rays, alpha particles, beta particles, gamma rays
- **D** alpha particles, beta particles, X-rays, gamma rays

45 For the following medical imaging techniques, list the radiation type that is utilised.
- **a** magnetic resonance imaging
- **b** single photon emission computed tomography
- **c** positron emission topography
- **d** computed tomography

46 A patient who is undergoing a course of radiotherapy is exposed to radiation. Their bladder, stomach and liver all absorb equal amounts of radiation. If the effective dose for this treatment is 8.8 mSv, what dose of radiation is absorbed by each of the organs?

47 I-123 can be used in the diagnosis of thyroid function. As part of the process of producing I-123, a Xe-124 atom is bombarded with a proton. This process also produces a daughter nucleus and two neutrons.
- **a** Write a balanced equation to depict this process.
- **b** The new element then decays by emitting a positron to form Xe-123. Write a balanced equation to depict this process.
- **c** And finally, Xe-123 decays to produce I-123 and another subatomic particle. Write a balanced equation to depict this process.

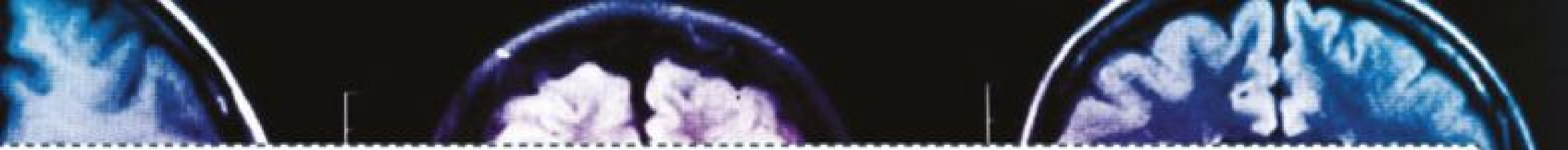
48 I-131 can be used to treat an overactive thyroid. I-131 has a half-life of 8 days. The iodine sample initially has an activity of 40 kBq.
- **a** What would the activity of this sample of I-131 be after 40 days?

A patient of mass 80 kg is injected with this I-131 sample.
- **b** Explain why the activity of the I-123 would be less than the amount calculated in part a, once it has been injected into a patient.
- **c** Assuming the patient absorbs 20 mJ of energy from this iodine treatment, calculate the patient's absorbed dose. Give your answer in μGy.

49 Outline the differences between somatic effects and genetic effects of ionising radiation. Discuss some of the symptoms of these effects.

Particle accelerators

50 10 eV is a measure of energy. Which of the following are equivalent to 10 eV?
- **A** The energy an electron gains accelerating across a potential difference of 10 V.
- **B** The energy that a proton gains accelerating across a potential difference of –10 V.
- **C** 1.6×10^{-18} J
- **D** All of the above.



51 Which of the following lists includes only subatomic particles that are classified as hadrons.

A protons and neutrons
B protons and electrons
C neutrons and electrons
D protons and neutrinos

52 What is the speed of an electron that has been accelerated across a potential difference of 5 kV? Take the mass of an electron to be 9.1×10^{-31} kg. Ignore any relativistic effects.

A 1.3×10^{6} m s^{-1}
B 4.2×10^{7} m s^{-1}
C 3.3×10^{15} m s^{-1}
D 1.0×10^{17} m s^{-1}

53 To what energy level can the LHC accelerate particles?

A 200 GeV
B 14 TeV
C 200 TeV
D 400 TeV

54 Which of the following lists particle accelerators in order of their date of invention and subsequent operation.

A cathode ray, tube, cyclotron, LHC, synchrotron
B cathode ray tube, synchrotron, cyclotron LHC
C cathode ray tube, cyclotron, synchrotron, LHC
D cyclotron, cathode ray tube, synchrotron, LHC

55 Draw a basic sketch of a synchrotron showing the position of the main components (as listed below) and outline the role of these components.

- linac
- booster ring
- beamline
- storage ring

56 The LHC, located in Europe has been involved in many key experiments since it commenced operation in 2008. Which of its initial experiments was responsible for detecting the Higgs boson, and what did the discovery of the Higgs boson mean to our understanding of the Standard Model?

57 What is synchrotron light, and how is it produced?

58 List some of the applications for which particle accelerators are being used.

59 Electrons in the booster ring of a synchrotron can be accelerated to energy levels of 3 GeV.

a Determine the speed of electrons with kinetic energy of 3 GeV.
Do not take into account any relativistic effects. Take the mass of the electron to be 9.1×10^{-31} kg.

b Explain what is 'wrong' with your answer to part a.

c Electrons that are accelerated to energy levels of 3 GeV reach speeds of 99.99999% of the speed of light. Explain this observation in terms of Einstein's mass and energy relationship ($E = mc^2$).

Sport

60 A ball is dropped from an initial height, H, and rebounds to a height h.
v_H is the speed of the ball just before it hits the ground and v_h its speed just after it leaves the ground.
E_k is the kinetic energy of the ball just before it hits the ground and E_k the kinetic energy just after it leaves the ground.
Which of the following is not an expression for the coefficient of restitution of the ball?

A $\frac{v_h}{v_H}$

B $\frac{h}{H}$

C $\frac{\sqrt{h}}{\sqrt{H}}$

D $\frac{\sqrt{E_k}}{\sqrt{E_k}}$

The following information relates to questions 61 and 62. A cricket ball, with a momentum of 5 kg m s^{-1} east is struck by a cricket bat. The ball moves away from the bat with a momentum of 3 kg m s^{-1} west.

61 What is the change in momentum of the cricket ball?

A 2 kg m s^{-1} east
B 2 kg m s^{-1} west
C 8 kg m s^{-1} east
D 8 kg m s^{-1} west

62 What is the change in momentum of the cricket bat?

A 2 kg m s^{-1} east
B 2 kg m s^{-1} west
C 8 kg m s^{-1} east
D 8 kg m s^{-1} west

63 What is the angular speed of a golf ball if the ball makes 2400 revolutions in one minute?

A 40 rad s^{-1}
B 40π rad s^{-1}
C 80π rad s^{-1}
D 2400π rad s^{-1}

64 Two balls are thrown into the air by two different basketball players. The balls are thrown at different angles to the horizontal, both from the same initial height. Both basketballs reach the same height. Which of the following statements is true?

A The initial velocity of the basketballs is the same and the time of flight for the basketballs is the same.
B The initial velocity of the basketballs is the same but the time of flight for the basketballs is different.
C Only the horizontal component of the initial velocity of the basketballs is the same and the time of flight for the basketballs is the same.
D Only the vertical component of the initial velocity of the basketballs is the same and the time of flight for the basketballs is the same.

UNIT 2 • Area of Study 2

65 The air resistance (drag) on a ball travelling at speed v is determined to be F. If the speed of the ball increases to $3v$, what is the new value of the air resistance?

- **A** $\frac{1}{3}F$
- **B** F
- **C** $3F$
- **D** $9F$

66 The air resistance (drag) on a semicircular parachute travelling at speed v is determined to be F. Which of the following would not cause the air resistance on the circular parachute to increase by a factor of four?

- **A** Increasing the area of the parachute by a factor of two.
- **B** Increasing the radius of the parachute by a factor of two.
- **C** Increasing the diameter of the parachute by a factor of two.
- **D** Increasing the area of the parachute by a factor of four.

67 Consider a tennis ball that has been hit by a tennis racquet and is undergoing projectile motion. Ignoring the effects of air resistance, which of the following statements about the ball's motion is not correct?

- **A** The horizontal component of the ball's velocity during its motion is constant.
- **B** The vertical component of the ball's velocity during its motion is constant.
- **C** The acceleration of the ball during its motion is constant.
- **D** The net force on the ball during its motion is constant.

68 A ball rebounds to only half of its initial drop height. Determine the coefficient of restitution of the ball on this drop surface.

69 In a game of snooker, the cue ball (m = 150 g) strikes a red ball (m = 150 g). Initially, the cue ball is travelling at 0.5 m s^{-1} east and the red ball is travelling towards the cue ball at 0.2 m s^{-1} west. After the collision, the cue ball recoils at 0.2 m s^{-1} west.

Determine the final velocity of the red ball, and show that this collision can be considered elastic.

70 Explain why most collisions can be considered to be inelastic collisions.

71 In a game of lawn bowls, a bowl (m = 1.40 kg) is released at a speed of 1.85 m s^{-1}. Assuming that the ball only experiences static friction (μ = 0.35), determine the distance the ball travels before coming to rest.

72 In a game of tunnel ball, a basketball needs to roll a distance of 8.0 m along the ground. At what frequency must the ball rotate at so that it can cover this distance in 4.0 s? Take the diameter of the ball to be 25 cm.

73 A physics student experiments with a simple pendulum and records the following results:

Length of pendulum (m)	Time for 10 swings (s)	Period (s)	Period^2 (s^2)
0.1	6.3		
0.2	9.0		
0.3	11.0		
0.4	12.7		
0.5	14.2		

a Complete the table by calculating the period (s) of the pendulum and then the period squared (s^2).
The relationship between the length of a pendulum and the period is given:

$$T = 2\pi\sqrt{\frac{l}{g}}$$

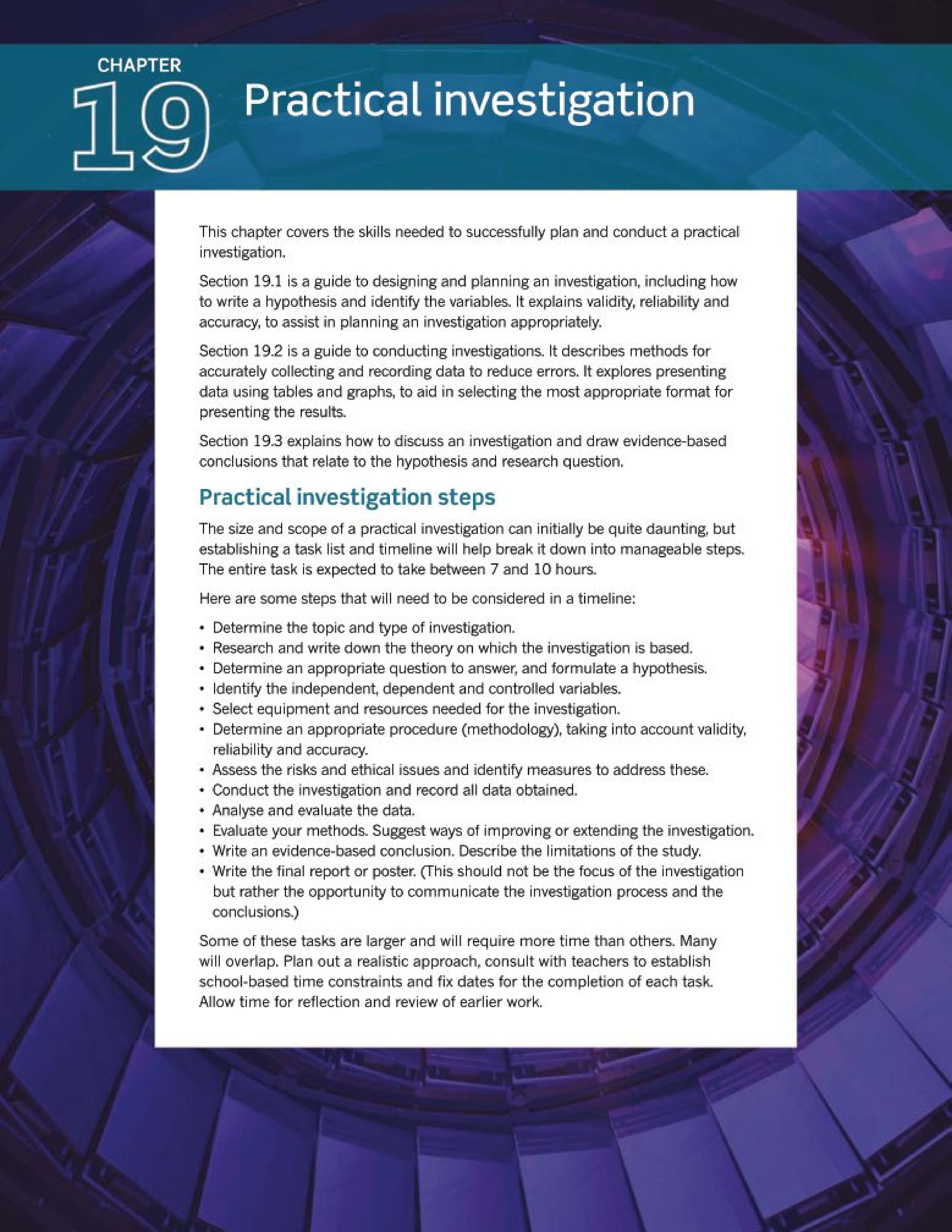
This can also be written as:

$$T^2 = 4\pi^2\frac{l}{g} \text{ or } T^2 = \frac{4\pi^2}{g}l$$

b Plot a graph of T^2 versus length and hence determine the value of g from your graph.

74 Draw a top view of a soccer ball travelling through the air to show how the Magnus force creates a sideways force on the soccer ball. Your diagram should clearly show the direction of spin of the ball, the direction of the velocity of the ball, the relative motion of the air to the ball and the direction of the Magnus force. Your diagram should also clearly indicate the areas of lower and higher pressure.

75 Compare the terminal velocities of two balls of identical diameter (15 cm) but with masses of 200 g and 400 g. The drag coefficient of both balls is 0.5 and the density of the air can be taken to be 1.25 kg m^{-3}. Use g = 9.8 m s^{-2}.



CHAPTER 19

Practical investigation

This chapter covers the skills needed to successfully plan and conduct a practical investigation.

Section 19.1 is a guide to designing and planning an investigation, including how to write a hypothesis and identify the variables. It explains validity, reliability and accuracy, to assist in planning an investigation appropriately.

Section 19.2 is a guide to conducting investigations. It describes methods for accurately collecting and recording data to reduce errors. It explores presenting data using tables and graphs, to aid in selecting the most appropriate format for presenting the results.

Section 19.3 explains how to discuss an investigation and draw evidence-based conclusions that relate to the hypothesis and research question.

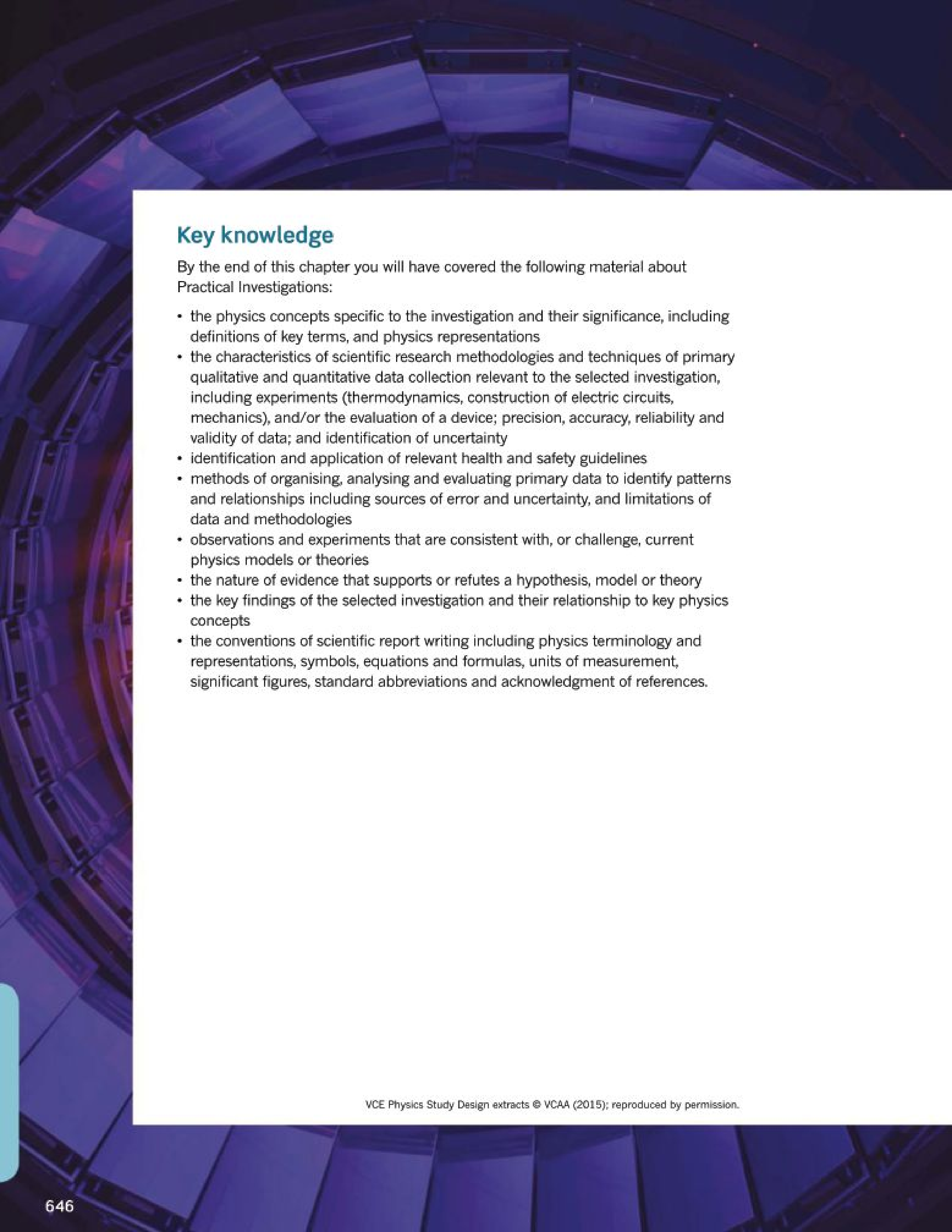
Practical investigation steps

The size and scope of a practical investigation can initially be quite daunting, but establishing a task list and timeline will help break it down into manageable steps. The entire task is expected to take between 7 and 10 hours.

Here are some steps that will need to be considered in a timeline:

- Determine the topic and type of investigation.
- Research and write down the theory on which the investigation is based.
- Determine an appropriate question to answer, and formulate a hypothesis.
- Identify the independent, dependent and controlled variables.
- Select equipment and resources needed for the investigation.
- Determine an appropriate procedure (methodology), taking into account validity, reliability and accuracy.
- Assess the risks and ethical issues and identify measures to address these.
- Conduct the investigation and record all data obtained.
- Analyse and evaluate the data.
- Evaluate your methods. Suggest ways of improving or extending the investigation.
- Write an evidence-based conclusion. Describe the limitations of the study.
- Write the final report or poster. (This should not be the focus of the investigation but rather the opportunity to communicate the investigation process and the conclusions.)

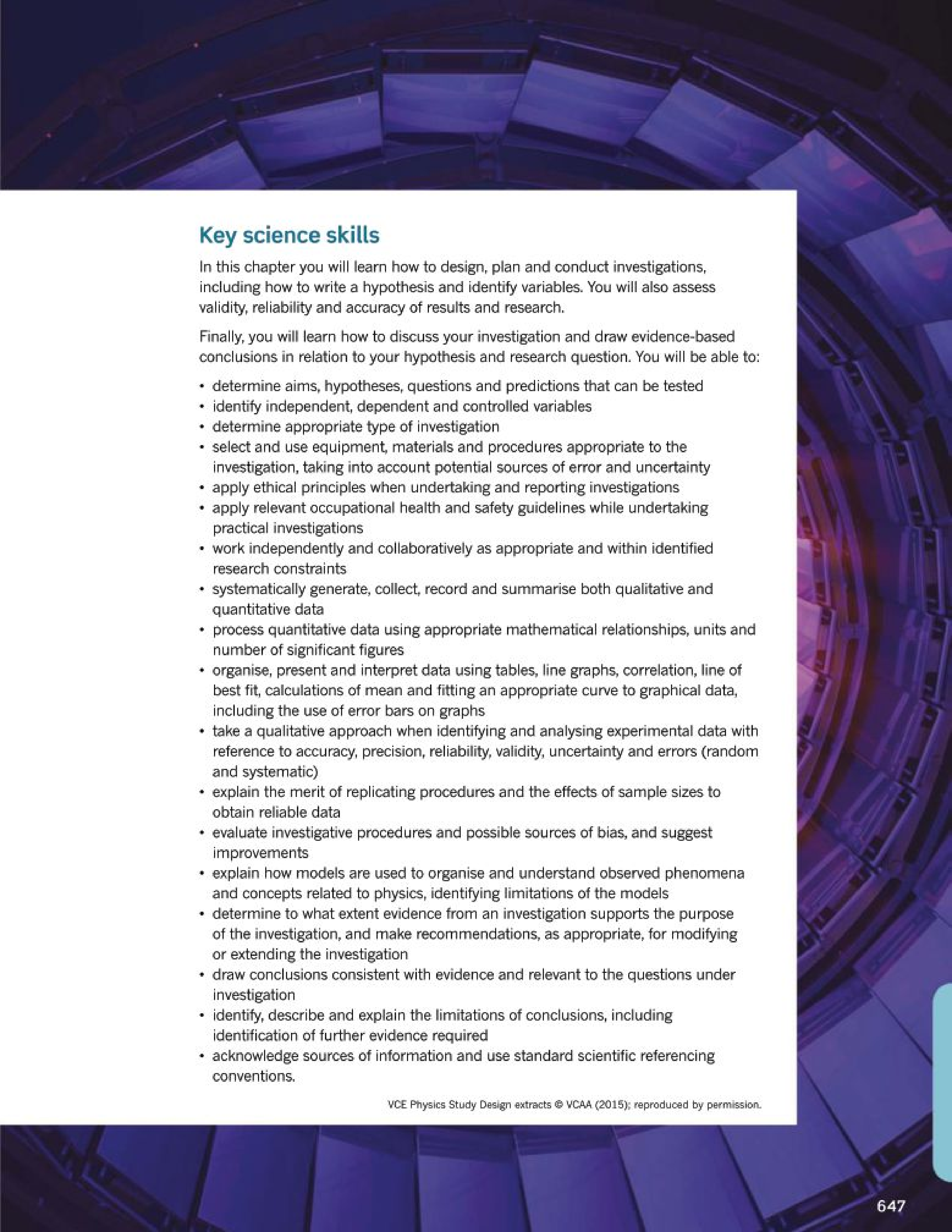
Some of these tasks are larger and will require more time than others. Many will overlap. Plan out a realistic approach, consult with teachers to establish school-based time constraints and fix dates for the completion of each task. Allow time for reflection and review of earlier work.



Key knowledge

By the end of this chapter you will have covered the following material about Practical Investigations:

- the physics concepts specific to the investigation and their significance, including definitions of key terms, and physics representations
- the characteristics of scientific research methodologies and techniques of primary qualitative and quantitative data collection relevant to the selected investigation, including experiments (thermodynamics, construction of electric circuits, mechanics), and/or the evaluation of a device; precision, accuracy, reliability and validity of data; and identification of uncertainty
- identification and application of relevant health and safety guidelines
- methods of organising, analysing and evaluating primary data to identify patterns and relationships including sources of error and uncertainty, and limitations of data and methodologies
- observations and experiments that are consistent with, or challenge, current physics models or theories
- the nature of evidence that supports or refutes a hypothesis, model or theory
- the key findings of the selected investigation and their relationship to key physics concepts
- the conventions of scientific report writing including physics terminology and representations, symbols, equations and formulas, units of measurement, significant figures, standard abbreviations and acknowledgment of references.



Key science skills

In this chapter you will learn how to design, plan and conduct investigations, including how to write a hypothesis and identify variables. You will also assess validity, reliability and accuracy of results and research.

Finally, you will learn how to discuss your investigation and draw evidence-based conclusions in relation to your hypothesis and research question. You will be able to:

- determine aims, hypotheses, questions and predictions that can be tested
- identify independent, dependent and controlled variables
- determine appropriate type of investigation
- select and use equipment, materials and procedures appropriate to the investigation, taking into account potential sources of error and uncertainty
- apply ethical principles when undertaking and reporting investigations
- apply relevant occupational health and safety guidelines while undertaking practical investigations
- work independently and collaboratively as appropriate and within identified research constraints
- systematically generate, collect, record and summarise both qualitative and quantitative data
- process quantitative data using appropriate mathematical relationships, units and number of significant figures
- organise, present and interpret data using tables, line graphs, correlation, line of best fit, calculations of mean and fitting an appropriate curve to graphical data, including the use of error bars on graphs
- take a qualitative approach when identifying and analysing experimental data with reference to accuracy, precision, reliability, validity, uncertainty and errors (random and systematic)
- explain the merit of replicating procedures and the effects of sample sizes to obtain reliable data
- evaluate investigative procedures and possible sources of bias, and suggest improvements
- explain how models are used to organise and understand observed phenomena and concepts related to physics, identifying limitations of the models
- determine to what extent evidence from an investigation supports the purpose of the investigation, and make recommendations, as appropriate, for modifying or extending the investigation
- draw conclusions consistent with evidence and relevant to the questions under investigation
- identify, describe and explain the limitations of conclusions, including identification of further evidence required
- acknowledge sources of information and use standard scientific referencing conventions.

VCE Physics Study Design extracts © VCAA (2015); reproduced by permission.

19.1 Designing and planning the investigation

Taking the time to carefully plan and design a practical investigation before beginning will help you to maintain a clear and concise focus throughout. Preparation is essential. Ensure you understand the theory behind the investigation and a prepare a detailed plan for the practical components of the investigation. This section is a guide to some of the key steps that should be taken when planning and designing a practical investigation.

FIGURE 19.1.1 There are many elements to a practical investigation, which may appear overwhelming to begin with. Taking a step-by-step approach will help the process and assist in completing a solid and worthwhile investigation.

DETERMINING THE TYPE OF INVESTIGATION

Several types of research methods are used in science. In this Area of Study (Unit 2, AoS 3) you are required to undertake an investigation related to an area covered in the course. Your teacher might suggest a particular topic or topics.

DEVELOPING AIMS AND QUESTIONS, FORMULATING HYPOTHESES AND MAKING PREDICTIONS

The research question, aim and hypothesis are interlinked. It is important to note that each of these can be refined as the planning of the investigation continues.

Research questions, aims and hypotheses

A research question is a statement defining what is being investigated. For example: What is the relationship between voltage and current?

An aim is a statement describing in detail what will be investigated. For example: The aim of the experiment is to investigate the relationship between voltage and the current in a circuit of constant resistance.

A hypothesis is a prediction based on previous knowledge and evidence or observations that attempts to answer the research question. For example: Increasing the voltage supplied to a circuit of constant resistance increases the current.

Formulating a question

Before formulating a question, it is good practice to conduct a literature review of the topic to be investigated. You should become familiar with the relevant scientific concepts and key terms.

During this review, write down questions or correlations as they arise.

Compile a list of possible ideas. Do not reject ideas that initially might seem impossible. Use these ideas to generate questions that are answerable.

Before constructing a hypothesis, formulate a question that needs an answer. This question will lead to a hypothesis when:

- the question is reduced to measurable **variables**
- a prediction is made based on knowledge and experience.

The question for the investigation must be related to a concept from Unit 2, which includes mechanics. Some examples of questions you might study include the following:

- How does the angle of release affect the projectile motion of an arrow?
- Does your body weight change when you travel up and down in a lift?

Evaluating your question

Once a question has been chosen, stop to evaluate the question before progressing. The question may need further refinement or even further investigation before it is suitable as a basis for an achievable and worthwhile investigation. A major planning point is to not attempt something that it is not possible to complete in the time available or with the resources on hand. It might be a little difficult to create a particularly complicated device with the facilities available in the school laboratory.

To evaluate the question, consider the following:

- Relevance: Is the question related to the appropriate area of study?
- Clarity and measurability: Can the question be framed as a clear hypothesis? If the question cannot be stated as a specific hypothesis, then it is going to be very difficult to complete the research.
- Time frame: Can the question be answered within a reasonable period of time? Is the question too broad?
- Knowledge and skills: Do you have a level of knowledge and a level of laboratory skills that will allow the question to be explored? Keep the question simple and achievable.
- Practicality: Are resources, such as laboratory equipment and materials, likely to be readily available? Keep things simple. Avoid investigations that require sophisticated or rare equipment. More-readily-available equipment includes timing devices, objects that could be used as projectiles, a tape measure and other common laboratory equipment.
- Safety and ethics: Consider the safety and ethical issues associated with the question you will be investigating. If there are issues, can these be addressed?
- Advice: Seek advice from the teacher about the question. Their input may prove very useful. Their experience may lead them to consider aspects of the question that you have not thought about.

Hypothesis

A hypothesis is a prediction that is based on evidence and prior knowledge. A hypothesis often takes the form of a proposed relationship between two or more variables in a cause and effect relationship; or in other words, 'If X is done, then Y will occur.'

Here are some examples of hypotheses:

- For a constant force, if the mass is increased the acceleration is decreased.
- If two objects are simultaneously dropped vertically from the same height they will both land at the same time.
- If the take-off angle is constant, the athlete who has the highest velocity at take-off will jump with the greatest horizontal displacement.
- A gymnast who has a set angular momentum when in the air will rotate faster during a somersault when they tuck their legs in towards their chest than if they keep their legs stretched out.

Variables

A good scientific hypothesis can be tested, that is supported or refuted, through investigation. To be a testable hypothesis, it should be possible to measure both what is changed or carried out and what will happen. The factors that are changed during an experiment or investigation are called the variables. An experiment or investigation determines the relationship between variables and measures the results.

There are three categories of variables:

- The **independent variable** is the variable that is controlled by the researcher (the one that is selected and changed).
- The **dependent variable** is the variable that may change in response to a change in the independent variable. This is the variable that will be measured or observed.
- **Controlled variables** are all the variables that must be kept constant during the investigation. Only test one variable at a time, otherwise it cannot be stated that the changes in the dependent variable are the result of changes in the independent variable.

Read the following example.

Hypothesis: If the take-off angle is constant, the athlete who has the highest velocity at take-off will jump with the greatest horizontal displacement.

- independent variable: take-off velocity
- dependent variable: horizontal jump displacement
- controlled variables: take-off angle, air resistance (including wind)

Completing a table like Table 19.1.1 will assist in evaluating the question or questions.

Research question	How does the angle of release of an arrow affect its projectile motion?
Independent variable	the angle of release of the arrow
Dependent variable	range of flight
Controlled variables	mass of the arrow, the extension of the bow (spring potential energy stored in the system), elevation of the arrow at release, height of release of the arrow, height at which arrow lands, release velocity of the arrow
Potential hypothesis	The range of flight is highest for a release angle of 45°.

TABLE 19.1.1 Break the question down to determine the variables.

Qualitative and quantitative variables

Variables are either qualitative or quantitative, with further subsets in each category.

- **Qualitative variables** can be observed but not measured. They can only be sorted into groups or categories such as brightness, type of material of construction, or type of device.
  - Nominal variables are categorical variables in which the order is not important; for example, the type of material or type of device.
  - Ordinal variables are categorical variables in which order is important and groups have an obvious ranking or level; for example, brightness (see Figure 19.1.2).
- **Quantitative variables** can be measured. Length, area, weight, temperature and cost are all examples of quantitative data.
  - Discrete variables consist of only integer numerical values, not fractions; for example, the number of pins in a packet, the number of springs connected together, or the energy levels in atoms.
  - Continuous variables allow for any numerical value within a given range; for example, the measurement of temperature, length, weight and frequency.

FIGURE 19.1.2 When recording qualitative data describe in detail how each variable will be defined. For example, if recording the brightness of light globes, pictures are a good way of clearly defining what each assigned term represents.

In this investigation, you must design and undertake an investigation involving two independent variables, one of which should be a continuous variable.

Formulating a hypothesis

Once the research question is confirmed, formulating a hypothesis comes next. A hypothesis requires a proposed relationship between two variables. It should predict that a relationship exists or does not exist.

Identify the two variables in your question. State the independent and dependent variables.

For example: If I do/change this (independent variable), then this (dependent variable) will happen.

A good hypothesis should:

- be a statement
- be based on information contained in the research question (purpose)
- be worded so that it can be tested in the experiment
- include an independent and a dependent variable
- include variables that are measurable.

The hypothesis should also be falsifiable. This means that a negative outcome would disprove it. For example, the hypothesis that all apples are round cannot be proved beyond doubt, but it can be disproved—in other words, it is falsifiable. In fact, only one square apple is needed to disprove this hypothesis. Unfalsifiable hypotheses cannot be proved by science. These include hypotheses on ethical, moral and other subjective judgements.

Defining the aim of the investigation

The aims are the key steps required to test the hypothesis. Each aim should directly relate to the variables in the hypothesis, and describe how each will be measured. The aims do not need to include the details of the method.

Example

- Hypothesis 1: When the force is kept constant, the acceleration decreases with increasing mass.
- Extension: When the force is kept constant, doubling the mass halves the acceleration.
- Hypothesis 2: When the mass is kept constant, the acceleration increases with increasing force.
- Extension: When the mass is kept constant, doubling the force doubles the acceleration.
- Aim: The aim of the experiment is to investigate the relationship between force, mass and acceleration.

In the first stage of the experiment, mass will be the independent variable (select a number of different masses) and the force is constant. The resulting acceleration (dependent variable) will be measured.

Then in the second stage of the experiment, force will be the independent variable (you select a number of different forces) and the mass will be kept constant. The resulting acceleration (dependent variable) will be measured.

These two investigations when combined create the classic Newton's second law experiment.

- Hypothesis 1 should give a result that mass is inversely proportional to the acceleration.
- Hypothesis 2 should give a result that force is proportional to the acceleration.

The combined result gives:

$F_{net} = ma$

Using the data collected from both stages of the experiment, the relationship between the three variables can be determined.

This level of 'neatness' is not always possible, especially with a student-designed experiment, but you should strive towards this.

Writing the methodology

The methodology of your investigation is a step-by-step procedure. When detailing the methodology, ensure it complies as a valid, reliable and accurate investigation.

Validity

Validity refers to whether an experiment or investigation is in fact testing the set hypothesis and aims. Is the investigation obtaining data that is relevant to the question? Is it flawed?

To ensure an investigation is valid, it should be designed so that only one variable is being changed at a time. The remaining variables must remain constant so that meaningful conclusions can be drawn about the effect of each variable in turn.

To ensure validity, carefully determine:

- the independent variable; that is, the variable that will be changed, and how it will change
- the dependent variable; that is, the variable that will be measured
- the controlled variables; that is, the variables that must remain constant, and how they will be maintained.

Reliability

Reliability refers to the notion that the experiment can be repeated many times and will obtain consistent results. Maintain the investigation's reliability by:

- defining the control
- ensuring there is sufficient replication of the experiment.

The control is an identical experiment carried out at the same time, except that in the control experiment the independent variable is not changed. A control can be:

- negative: the effect or change is expected in the experimental group but not in the control group
- positive: the effect or change is expected in the control group but not in the experimental group.

The expectations are based on previous experiments or observations. When the controls do not behave as expected, the data obtained from an experiment or observation is not reliable.

It is also important to determine how many times the experiment needs to be replicated (see Figure 19.1.3). Many scientific investigations lack sufficient repetition to ensure that the results can be considered reliable and repeatable.

FIGURE 19.1.3 Replication increases the reliability of your investigation. It ensures that if anyone repeats the investigation they will obtain similar data.

- Repeat readings: repeat each reading three times, record each measurement and then average the three measurements. This reduces systematic errors and allows random errors to be identified. If a reading differs too much from the rest (known as an outlier), discard it before averaging.
- Sample size: where there might be differences in construction or manufacture of a sample, then there should be various samples with the same conditions in the same experiment. The greater the sample size the more reliable the data.
- Repeats: if possible repeat the experiment on a different day. Don't change anything. If the results are not the same, think about what could have happened. For example, was the equipment faulty, or were all the controlled variables correctly identified. Repeat the experiment a third time to confirm which run was correct. More repeats is better; three is a good number but, if time and resources allow, aim for five.

Accuracy and precision

Precision refers to the minimum difference the instrument can measure; for example, units and decimal places. Accuracy refers to the ability to obtain the correct measurement.

Are the instruments to be used sensitive enough? What units will be used? Build some testing into your investigation to confirm the accuracy and reliability of the equipment and your ability to read the information obtained.

Reasonable steps to ensure the accuracy of the investigation include considering:

- the unit in which the independent and dependent variables will be measured
- the instrument that will be used to measure the independent and dependent variables.

Select and use appropriate equipment, materials and methods. For example, select equipment that measures to smaller degrees to reduce uncertainty and repeat the measurements to confirm them.

Describe the materials and method in appropriate detail in the logbook. This should ensure that every measurement can be repeated and the same result obtained within reasonable margins of experimental error (less than 5% is reasonable).

Data analysis

Data analysis is part of the method. Consider how the data will be presented and analysed. A wide range of analysis tools are available. For example, tables organise data so that patterns can be established and graphs can show relationships and comparisons. In fact, preparing an empty table showing the data that needs to be obtained will help in the planning of the investigation.

The nature of the data being collected, such as whether the variables are qualitative or quantitative, influences the type of method or tool that you can use to analyse the data. The aims and the hypothesis will also influence the choice of analysis tool.

Modifying the methodology

The methodology may need modifying as the investigation is carried out. The following actions will help to determine any issues in the methodology and how to modify them:

- Record everything.
- Be prepared to make changes to the approach.
- Note any difficulties encountered and the ways they were overcome. What were the failures and successes? Every test carried out can contribute to the understanding of the investigation as a whole, no matter how much of a disaster it may first appear.
- Do not panic. Go over the theory again, and talk to the teacher and other students. A different perspective can lead to a solution.

If the expected data is not obtained, don't worry. As long as it can be critically and objectively evaluated, the limitations of the investigation are identified and further investigations proposed, the work is worthwhile.

COMPLYING WITH ETHICAL AND SAFETY GUIDELINES

Ethical considerations

Some investigations require an ethics approval; consult with the teacher. In fact, when deciding on an investigation, identify all possible ethical considerations and evaluate their necessity or ways that can reduce or mitigate them.

Occupational health and safety

While planning for an investigation, it is important for the safety of yourself and the safety of others that the potential risks are considered.

Everything we do has some risk involved. Risk assessments are performed to identify, assess and control hazards. A risk assessment should be performed for any situation, in the laboratory or outside in the field. Always identify the risks and control them to keep everyone safe. For example, carry out voltage–current experiments with low voltages (less than 6.0 V DC or 4 × 1.5 V batteries) coupled to resistors so that the currents in the circuits are of the order of milliamps. AT ALL TIMES avoid direct exposure to 240 V AC household voltages (see Figure 19.1.4).

FIGURE 19.1.4 When planning an investigation you need to identify, assess and control hazards.

To identify risks think about:

- the activity that will be carried out
- the equipment or chemicals that will be used.

The following hierarchy of risk controls is organised from most effective to least effective:

1. *Elimination:* Eliminate dangerous equipment, procedures or substances.
2. *Substitution:* Find different equipment, procedures or substances to use that will achieve the same result, but have less risk associated.
3. *Isolation:* Ensure there is a barrier between the person and the hazard. Examples include physical barriers such as guards in machines, or fume hoods to work with volatile substances.
4. *Engineering controls:* Modify equipment to reduce risks.
5. *Administrative controls:* Provide guidelines, special procedures, warning signs and safe behaviours for any participants.
6. *Personal protective equipment (PPE):* Wear safety glasses, lab coats, gloves and respirators etc. where appropriate, and provide these to other participants.

Science outdoors

Sometimes investigations and experiments will be carried out outdoors. Working outdoors has its own set of potential risks and it is equally important to consider ways of eliminating or reducing these risks.

As an example, read Table 19.1.2, which contains examples of risks associated with field work in a national park.

Risks	Control measures
sunburn	wear sunscreen, a hat and sunglasses
hot or cold weather	wear clothing to protect against heat or cold
projectile launch	create barriers so that people know not to enter the area
trip hazards	minimise the use of cables (electrical, computer) and cover them up with matting be aware of tree roots, rocks etc.

TABLE 19.1.2 Risks associated with fieldwork in a national park.

First aid measures

Minimising the risk of injury reduces the chance of requiring first aid assistance. However, it is still important to have someone with first aid training present during practical investigations. Always tell the teacher or laboratory technician if an injury or accident happens.

Personal protective equipment

Everyone who works in a laboratory wears items that help keep them safe. This is called **personal protective equipment (PPE)** and includes:

- safety glasses
- shoes with covered tops
- disposable gloves when handling chemicals
- a disposable apron or a lab coat if there is risk of damage to clothing
- ear protection if there is risk to hearing.

19.1 Review

SUMMARY

- An aim is a statement describing in detail what will be investigated. For example: The aim of the experiment is to investigate the relationship between force, mass and acceleration.
- A hypothesis is a prediction based on previous knowledge and evidence or observations that attempts to answer the research question. For example: With the force kept constant, the acceleration decreases with increasing mass.
- Once a question has been chosen, stop to evaluate the question before progressing. The question may need further refinement or even further investigation before it is suitable as a basis for an achievable and worthwhile investigation. A major planning point is to not attempt something that it is not possible to complete in the time available or with the resources on hand. It might be a little difficult to create a particularly complicated device with the facilities available in the school laboratory.
- There are three categories of variables:
  - The independent variable is the variable that is controlled by the researcher (the one that is selected and changed).
  - The dependent variable is the variable that may change in response to a change in the independent variable. This is the variable that will be measured or observed.
  - Controlled variables are all the variables that must be kept constant during the investigation. Only test one variable at a time. Otherwise, it cannot be stated that the changes in the dependent variable are the result of changes in the independent variable.
- The methodology of your investigation is a step-by-step procedure. When detailing the methodology, ensure it complies as a valid, reliable and accurate investigation.
- It is also important to determine how many times the experiment needs to be replicated. Many scientific investigations lack sufficient repetition to ensure that the results can be considered reliable and repeatable.
- Data analysis is part of the method. Consider how the data will be presented and analysed. A wide range of analysis tools could be used. For example, tables organise data so that patterns can be established and graphs can show relationships and comparisons.

KEY QUESTIONS

1 In a practical investigation the student changes the voltage by adding or subtracting batteries in series to the circuit.
   a How could the voltage be a discrete value?
   b How could it be continuous?

2 In another experiment the student uses the following range of values to describe the brightness of a light: dazzling, bright, glowing, dim, off
   What type of measurement is the variable 'brightness'?

3 Select the best hypothesis from the three options below. Give reasons for your choice.
   A Hypothesis 1: Take-off angular momentum and inertia affect angular (rotational) velocity.
   B Hypothesis 2: Body position during angular airborne motion affects its inertia.
   C Hypothesis 3: A springboard diver's angular (rotational) velocity is slower when they hold a stretched (layout) position than when they are in a tuck position, if they take off with the same angular momentum.

4 Give the correct term that describes an experiment with each of the following conditions.
   a The experiment addresses the hypothesis and aims.
   b The experiment is repeated and consistent results are obtained.
   c Appropriate equipment is chosen for the desired measurements.

5 A student wanted to find out if you can hit a ball harder with a two-handed grip of the bat instead of a one-handed grip. What would be the independent variable for their experiment?

19.2 Conducting investigations and recording and presenting data

FIGURE 19.2.1 When carrying out your investigation try to maintain high standards to minimise potential errors.

Once the planning and design of a practical investigation is complete, the next step is to undertake the investigation and record the results. As with the planning stages, there are key steps and skills to keep in mind to maintain high standards and minimise potential error throughout the investigation (Figure 19.2.1).

This section will focus on the best methods of conducting a practical investigation, of systematically generating, recording and processing data and of presenting it in a concise and clear manner.

CONDUCTING INVESTIGATIONS TO COLLECT AND RECORD DATA

For an investigation to be scientific, it must be objective and systematic. Ensuring familiarity with the methodology and protocols before beginning will help you to achieve this.

When working, keep asking questions. Is the work biased in any way? If changes are made, how will they affect the study? Will the investigation still be valid for the aim and hypothesis?

It is essential that during the investigation the following are recorded in the logbook:

- all quantitative and qualitative data collected
- the methods used to collect the data
- any incident, feature or unexpected event that may have affected the quality or validity of the data.

The data recorded in the logbook is the **raw data**. Usually this data needs to be processed in some manner before it can be presented. If an error occurs in the processing of the data or you decide to present the data in an alternative format, the recorded raw data will always be available for you to refer back to.

IDENTIFYING ERRORS

Most practical investigations have errors associated with them. Errors can occur for a variety of reasons. Being aware of potential errors helps you to avoid or minimise them. For an investigation to be accurate, it is important to identify and record any errors.

There are two types of errors:

- systematic errors
- random errors.

FIGURE 19.2.2 To avoid a systematic error, make sure that you are using measuring equipment correctly. Laser speed guns, for example, need to be placed on a stationary support so the aim point is held on a single target point for the duration of the read.

Systematic errors

A **systematic error** is an error that is consistent and will occur again if the investigation is repeated in the same way.

Systematic errors are usually a result of instruments that are not calibrated correctly or methods that are flawed.

An example of a systematic error would be if a ruler mark indicating 5 cm from 0 cm was actually only 4.9 cm from 0 cm due to a manufacturing error or shrinkage of the wood. Another example would be if the researcher repeatedly used a piece of equipment incorrectly throughout the entire investigation. Figure 19.2.2 shows how traffic police reduce systematic errors in their data collection.

Random errors

Random errors occur in an unpredictable manner and are generally small. A random error could be, for example, the result of a researcher reading the same result correctly one time and incorrectly another time. Another example would be if an instrument were affected by a power cut or low battery power.

Techniques for reducing error

Designing the method carefully, including selection and use of equipment, will help reduce errors.

Appropriate equipment

Use the equipment that is best suited to the data that needs to be collected to validate the hypothesis. Determining the units of the data being collected and at what scale will help to select the correct equipment. Using the right unit and scale will ensure that measurements are more accurate and precise (with smaller systematic errors).

Significant figures are the numbers that convey meaning and precision. The number of significant figures used depends on the scale of the instrument. It is important to record data to the number of significant figures available from the equipment or observation. Using either a greater or smaller number of significant figures can be misleading.

Review the following examples to learn more about significant figures:

- 15 has two significant figures
- 3.5 has two significant figures
- 3.50 has three significant figures
- 0.037 has two significant figures
- 1401 has four significant figures.

To calculate gravitational potential energy (E_g), the formula is $E_g = mg\Delta h$. If $g = 9.81\ \text{m s}^{-2}$, mass = 7.50 kg, height = 0.64 m (64 cm):

$E_g = 9.81 \times 7.50 \times 0.64 = 47.09\ \text{J}$

But only quote the answer to the least number of significant figures in the data; that is, to two significant figures, so $E_g = 47\ \text{J}$.

Although digital scales can measure to many more than two figures and calculators can give 12 figures, be sensible and follow the significant figure rules.

Calibrated equipment

Some equipment, such as some motion sensors, needs to be calibrated before use to account for the temperature at the time. Before carrying out the investigation, make sure the instruments or measuring devices are properly calibrated and are, in general, functioning correctly. For example, measure the temperature and apply a correction to the speed of sound to calibrate a motion sensor if necessary.

Correct use of equipment

Use the equipment properly. Ensure any necessary training has been done to use the equipment and that you have had an opportunity to practice using the equipment before beginning the investigation. Improper use of equipment can result in inaccurate, imprecise data with large errors, and the validity of the data can be compromised.

Incorrect reading of measurements is a common misuse of equipment. Make sure all of the equipment needed in the investigation can be used correctly and record the instructions in detail so they can be referred back to if the data doesn't appear correct.

RECORDING AND PRESENTING QUANTITATIVE DATA

Raw data is unlikely to be used directly to validate the hypothesis. However, raw data is essential to the investigation and plans for collecting the raw data should be made carefully. Consider the formulas or graphs that will be used to analyse the data at the end of the investigation. This will help to determine the type of raw data that needs to be collected in order to validate the hypothesis.

For example, to calculate take-off velocity for a vertical jump, three sets of raw data will need to be collected using a force platform: the athlete's standing body weight, the ground reaction force and the time during the vertical jump. The data can then be processed to obtain the take-off impulse.

Once you have determined the data that needs to be collected, prepare a table in which to record the data.

ANALYSING AND PRESENTING DATA

The raw data that has been obtained needs to be presented in a way that is clear, concise and accurate.

There are a number of ways of presenting data, including tables, graphs, flow charts and diagrams. The best way of visualising the data depends on its nature. Try several formats before making a final decision, to create the best possible presentation.

Presenting raw and processed data in tables

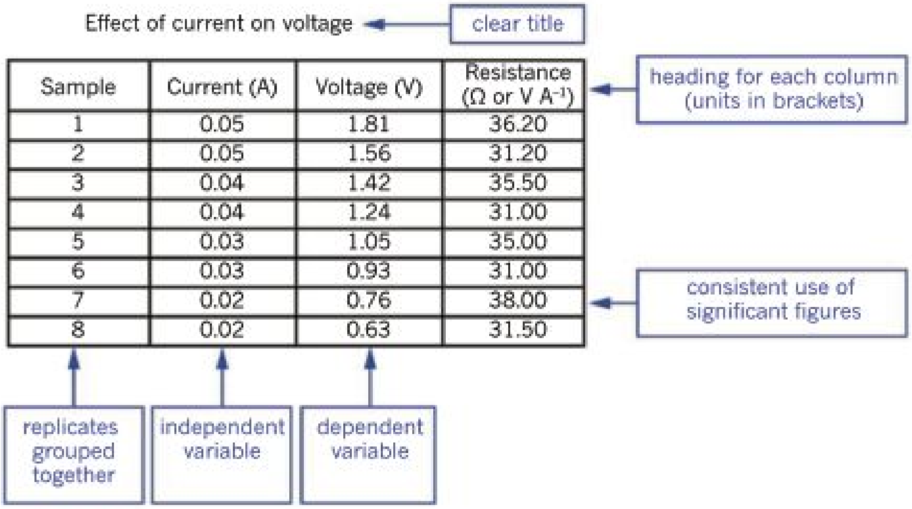
Tables organise data into rows and columns and can vary in complexity according to the nature of the data. Tables can be used to organise raw data and processed data or to summarise results.

The simplest form of a table is a two-column format. In a two-column table, the first column should contain the independent variable (the one being changed) and the second column should contain the dependent variable (the one that may change in response to a change in the independent variable).

Tables should have the following features:

- a descriptive title
- column headings (including the unit)
- aligned figures (align the decimal points)
- the independent variable placed in the left column
- the dependent variable placed in the right column.

Look at the table in Figure 19.2.3, which has been used to organise raw and processed data about the effect of current on voltage.



Effect of current on voltage

Sample	Current (A)	Voltage (V)	Resistance (Ω or $V A^{-1}$)
1	0.05	1.81	36.20
2	0.05	1.56	31.20
3	0.04	1.42	35.50
4	0.04	1.24	31.00
5	0.03	1.05	35.00
6	0.03	0.93	31.00
7	0.02	0.76	38.00
8	0.02	0.63	31.50

FIGURE 19.2.3 A simple table listing the raw data obtained in the second and third columns and processed data in the fourth column.

A table of processed data usually presents the average values of replicates, the **mean**. However, the mean on its own does not provide an accurate picture of the results.

To report processed data more accurately, the **uncertainty** should be presented as well.

Uncertainty

When there is a range of measurements of a particular value, the mean must be accompanied by the uncertainty, for your results to be presented as a mean in an accurate way. In other words, the mean must be accompanied by a description of the range of data obtained.

Uncertainty is calculated by:

uncertainty = ± (maximum variance from the mean)

For example, the speed, in km h^{-1}, of cars travelling down a certain road was:

46, 50, 55, 48, 50, 58, 45

The average speed would be:

$(46 + 50 + 55 + 48 + 50 + 58 + 45) \div 7 = 50$ km h^{-1}

The uncertainty would be the maximum variance from the average: 58 is 8 above the average, so the uncertainty is 8.

This data should be presented as:

Average speed is 50 ± 8 km h^{-1}.

Other descriptive statistics measures

The mean and the uncertainty are statistical measures that help describe data accurately. Other statistical measures that can be used, depending on the data obtained, are:

- **mode**: the mode is the value that appears most often in a data set. This measure is useful to describe qualitative or discrete data (for example, the mode of the values 0.01, 0.01, 0.02, 0.02, 0.02, 0.03, 0.04 is 0.02).
- **median**: the median is the 'middle' value of an ordered list of values (for example, the median of the values 5, 5, 8, 8, 9, 10, 20 is 8). The median is used when the data range is spread, for example, due to the presence of unusual results, making the mean unreliable.

Graphs

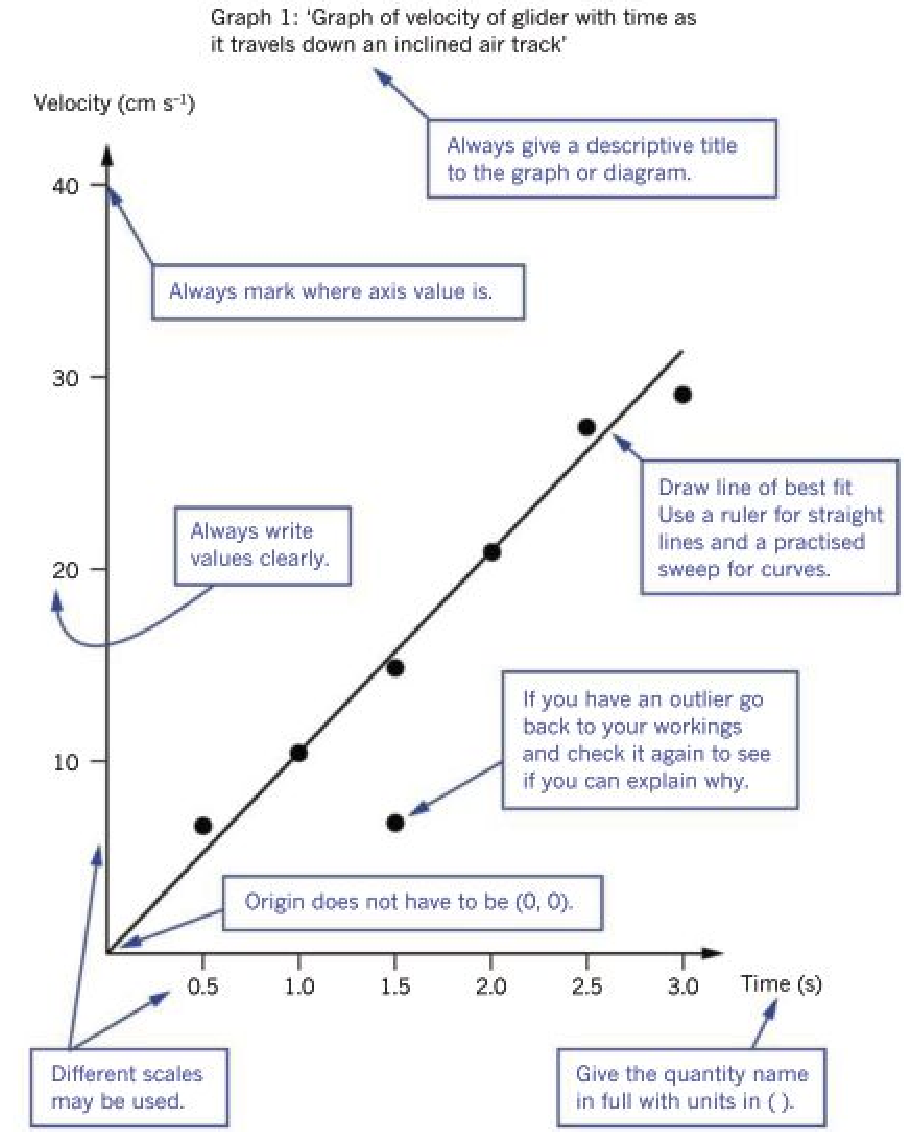
In general, tables provide more detailed data than graphs, but it is easier to observe trends and patterns in data in graphical form than in tabular form.

Graphs are used when two variables are being considered and one variable is dependent on the other. The graph shows the relationship between the variables.

There are several types of graphs that can be used, including line graphs, bar graphs and pie charts. The best one to use will depend on the nature of the data.

General rules to follow when making a graph (see Figure 19.2.4) include the following:

- Keep the graph simple and uncluttered.
- Use a descriptive title.
- Represent the independent variable on the *x*-axis and the dependent variable on the *y*-axis.
- Make axes proportionate to the data.
- Clearly label axes with both the variable and the unit in which it is measured.



Graph 1: 'Graph of velocity of glider with time as it travels down an inclined air track'

Always give a descriptive title to the graph or diagram.

Always mark where axis value is.

Always write values clearly.

Draw line of best fit Use a ruler for straight lines and a practised sweep for curves.

If you have an outlier go back to your workings and check it again to see if you can explain why.

Origin does not have to be (0, 0).

Different scales may be used.

Give the quantity name in full with units in ( ).

FIGURE 19.2.4 A graph shows the relationship between two variables.

Line graphs

Line graphs are a good way of representing continuous quantitative data. In a line graph, the values are plotted as a series of points on the graph. There are two ways of joining these points:

- A line can be ruled from each point to the next (see Figure 19.2.5(a)). It shows the overall trend; it is not meant to predict the value of the points between the plotted data.
- The points can be joined with a single smooth straight or curved line (see Figure 19.2.5(b)). This creates a trend line, also known as a line of best fit. The line of best fit does not have to pass through every point but should go close to as many points as possible. It is used when there is an obvious trend between the variables.

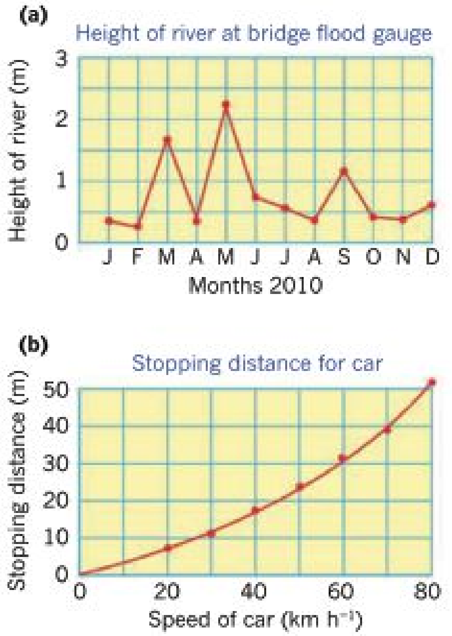


FIGURE 19.2.5 (a) The data in the graph is joined from point to point. (b) The data in graph is joined with a line of best fit, which shows the general trend.

Outliers

Sometimes when the data is collected, there may be one point that does not fit with the trend and is clearly an error. This is called an **outlier**. An outlier is often caused by a mistake made in measuring or recording data, or from a random error in the measuring equipment. If there is an outlier, include it on the graph, but ignore it when adding a line of best fit (as in Figure 19.2.4 where the point (1.5, 6) is an outlier).

19.2 Review

SUMMARY

- It is essential that during the investigation, the following are recorded in the logbook:
  - all quantitative and qualitative data collected
  - the methods used to collect the data
  - any incident, feature or unexpected event that may have affected the quality or validity of the data.
- A systematic error is an error that is consistent and will occur again if the investigation is repeated in the same way. Systematic errors are usually a result of instruments that are not calibrated correctly or methods that are flawed.
- Random errors occur in an unpredictable manner and are generally small. A random error could be, for example, the result of a researcher reading the same result correctly one time and incorrectly another time.
- The number of significant figures used depends on the scale of the instrument used. It is important to record data to the number of significant figures available from the equipment or observation.
- The simplest form of a table is a two-column format in which the first column contains the independent variable (the one being changed) and the second column contains the dependent variable (the one that may change in response to a change in the independent variable).
- When there is a range of measurements of a particular value, the mean must be accompanied by the uncertainty, for your results to be presented as a mean in an accurate way.
- General rules to follow when making a graph include the following:
  - Keep the graph simple and uncluttered.
  - Use a descriptive title.
  - Represent the independent variable on the *x*-axis and the dependent variable on the *y*-axis.
  - Make axes proportionate to the data.
  - Clearly label axes with both the variable and the unit in which it is measured.

KEY QUESTIONS

1 The masses of 1 cm^3 cubes of potato were recorded and the cubes placed in distilled water. After 60 minutes, the cubes were weighed again and the difference in mass was calculated. What type of error is involved:

a if the electronic scales only measured in 1 g increments?

b if the electronic scales were affected briefly by a power surge?

2 If using the quantities mass = 7.50 kg and speed = 1.4 m s^{-1} in a calculation, what would be the appropriate number of significant figures in the answer?

3 For the data set 21, 28, 19, 19, 25, 24 determine:

a the mean

b the mode

c the median.

4 Plot the following data set, assigning each variable to the appropriate axis on the graph.

Current (A)	Voltage (V)
0.06	2.07
0.05	1.56
0.04	1.24
0.03	0.93
0.02	0.63

5 How can the general pattern (trend) of a graph be represented once the points are plotted?

19.3 Discussing investigations and drawing evidence-based conclusions

Now that the chosen topic has been thoroughly researched and the investigation has been conducted and data collected, it is time to draw it all together. The final part of the investigation involves summarising the findings in an objective, clear and concise manner.

FIGURE 19.3.1 To discuss and conclude your investigation, utilise the raw and processed data.

EXPLAINING RESULTS IN THE DISCUSSION

The discussion is the part of the investigation where the evaluation and explanation of the investigation methods and results takes place. It is the interpretation of what the results mean.

The key sections of the discussion are:

- analysing and evaluating data
- evaluating the investigative method
- explaining the link between the investigation findings and the relevant physics concepts.

Consider the message to be conveyed to the audience, when writing the discussion. Statements need to be clear and concise. At the conclusion of the discussion, the audience must have a clear idea of the context, results and implications of the investigation.

ANALYSING AND EVALUATING DATA

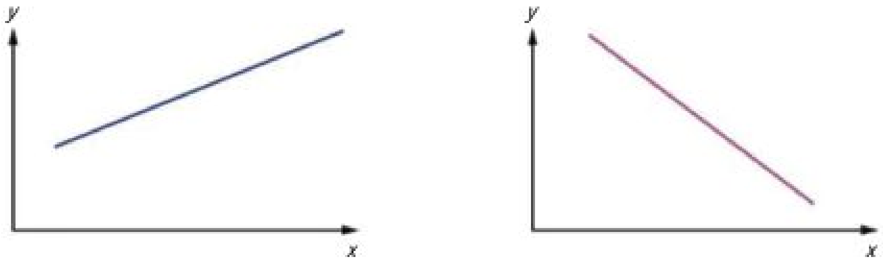
In the discussion, the findings of the investigation need to be analysed and interpreted.

- State whether a pattern, trend or relationship was observed between the independent and dependent variables. Describe what kind of pattern it was and specify under what conditions it was observed.
- Were there discrepancies, deviations or anomalies in the data? If so, these should be acknowledged and explained.
- Identify any limitations in the data you have collected. Perhaps a larger sample or further variations in the independent variable would lead to a stronger conclusion.

Trends in line graphs

Graphs are drawn to show the relationship, or trend, between two variables, as shown in Figure 19.3.2.

- Variables that change in linear or direct proportion to each other produce a straight, sloping trend line.
- Variables that change exponentially in proportion to each other produce a curved trend line.
- When there is an inverse relationship, one variable increases as the other variable decreases.
- When there is no relationship between two variables, one variable will not change even if the other changes.

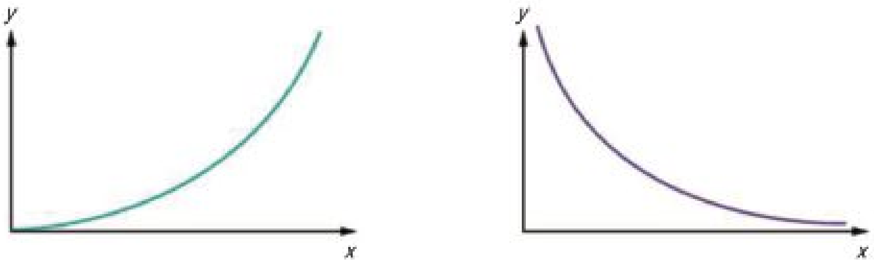


Direct or linear proportional relationship

- Variables change at the same rate (graph line is straight, slope is constant).
- Positive relationship—as x increases, y increases.

Inverse direct or linear proportional relationship

- Variables change at the same rate (graph line is straight, slope is constant).
- Negative relationship—as x increases, y decreases.

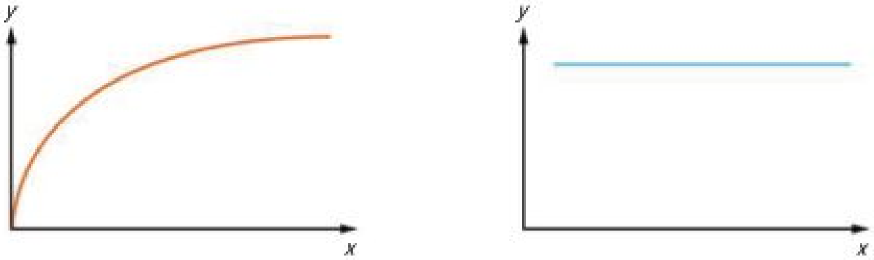


Exponential relationship

- As x increases, y increases slowly, then more rapidly.

Inverse exponential relationship

- As x increases, y decreases rapidly, then more slowly, until a minimum y value is reached.



Exponential rise, then levels off or plateaus (stops rising)

- As x increases, y increases rapidly at first, then slows, then finally does not increase at all—y reaches a maximum value.

No relationship between x and y

- As x increases, y remains the same.

FIGURE 19.3.2 Various relationships can exist between two variables.

Remember that the results may be unexpected. This does not make the investigation a failure. However, the findings must be related to the hypothesis, aims and method.

EVALUATING THE METHOD

It is important to discuss the limitations of the investigation method. Evaluate the method and identify any issues that could have affected the validity, accuracy, precision or reliability of the data. Sources of errors and uncertainty must also be stated in the discussion.

Once any limitations or problems in the methodology have been identified, recommend improvements on how the investigation could be conducted if repeated; for example, suggest how bias could be minimised or eliminated.

Bias

Bias may occur in any part of the investigation method, including sampling and measurements.

Bias is a form of systematic error resulting from the researcher's personal preferences or motivations. There are many types of bias, including:

- poor definitions of both concepts and variables (for example, classifying cricket pitch surfaces and their interaction with the ball according to resilience without defining 'slow' and 'fast')
- incorrect assumptions (for example, that footwear type, model and manufacturer does not affect ground reaction forces, and as a result failing to control this variable during an investigation on slip risk on different indoor and outdoor surfaces)
- errors in the investigation design and methodology (for example, taking a sample of a particular group of athletes that samples one particular gender more than the other in the group).

Some biases cannot be eliminated, but should at least be addressed in the discussion.

Accuracy and precision

In the discussion, evaluate the degree of accuracy and precision of the measurements for each variable of the hypothesis. Comment on the uncertainties obtained.

When relevant, compare the chosen method with any other methods that might have been selected, evaluating the advantages and disadvantages of the selected method and the effect on the results.

Reliability

When discussing the results, indicate the range of the data obtained from replicates. Explain how the sample size was selected. Larger samples are usually more reliable, but time and resources might have been scarce. Discuss whether the results of the investigation have been limited by the sample size.

The control group is important to the reliability of the investigation. A control group helps determine if a variable that should have been controlled has been overlooked and may explain any unexpected results.

Error

Discuss any source of systematic or random error and suggest ways of improving the investigation.

FIGURE 19.3.3 Honest evaluation and reflection play important roles in analysing methodology.

DISCUSSING RELEVANT PHYSICS CONCEPTS

To make the investigation more meaningful, it should be explained within the right context, meaning the related physics ideas, concepts, theories and models. Within this context, explain the basis for the hypothesis.

For example, if studying the impact of temperature on linear strain of a material (e.g. a rubber band), some of the contextual information to include in the discussion could be:

- the definition of linear strain
- the functions of linear strain
- the relationship between linear strain and temperature
- the definitions of material behaviour such as plastic and elastic
- the factors known to affect linear strain
- existing knowledge on the role of temperature on linear strain
- the ranges of temperatures investigated and the reason these temperatures were chosen
- the materials studied and the reasons for this choice
- methods of measuring the linear strain of a material.

Relating your findings to a physics concept

Once a context is established, a framework will have been created in which to discuss whether the data supported or refuted the hypothesis. Ask questions such as:

- Was the hypothesis supported?
- Has the hypothesis been fully answered? (If not, give an explanation of why this is so and suggest what could be done to either improve or complement the investigation.)
- Do the results contradict the hypothesis? If so, why? (The explanation must be plausible and must be based on the results and previous evidence.)

Providing a theoretical context also enables comparison of the results with existing relevant research and knowledge. After identifying the major findings of the investigation, ask questions such as:

- How does the data fit with the literature?
- Does the data contradict the literature?
- Do the findings fill a gap in the literature?
- Do the findings lead to further questions?
- Can the findings be extended to another situation?

Be sure to discuss the broader implications of the findings. Implications are the bigger picture. Outlining them for the audience is an important part of the investigation. Ask questions such as:

- Do the findings contribute to or impact on the existing literature and knowledge of the topic?
- Are there any practical applications for the findings?

DRAWING EVIDENCE-BASED CONCLUSIONS

A conclusion is usually a paragraph that links the collected evidence to the hypothesis and provides a justified response to the research question.

Indicate whether the hypothesis was supported or refuted and the evidence on which this is based (that is, the results). Do not provide irrelevant information. Only refer to the specifics of the hypothesis and the research question and do not make generalisations.

Read the examples of conclusions for the following hypothesis and research question.

Hypothesis: An increase in temperature will cause an increase in linear deformation (change in length) before failure.

- Poor response to the hypothesis: Linear deformation has value y_1 at temperature 1 and value y_2 at temperature 2.
- Better response to the hypothesis: An increase in temperature from 1 to 2 produces an increase in linear deformation of z in the rubber band.

Research question: Does temperature affect the maximum linear deformation the material can withstand?

- Poor response to the research question: The results show that temperature does affect the maximum deformation of a material.
- Better response to the research question: Analysis of the results of the effect of an increase in temperature from 1 to 2 in the rubber band support current knowledge on the effect that an increase in temperature has on increasing maximum linear deformation.

REFERENCES AND ACKNOWLEDGEMENTS

All the quotations, documents, publications and ideas used in the investigation need to be acknowledged in the 'references and acknowledgments' section in order to avoid plagiarism and to ensure authors are credited for their work. References and acknowledgements also give credibility to the study and allow the audience to locate information sources should they wish to study it further.

When referencing a book include in this order:

- author's surname and initials
- date of publication
- title
- publisher's name
- place of publication.

For example: Rickard G. et al. (2005), *Science Dimensions 1*, Pearson Education, Melbourne, Victoria.

When referencing a website include in this order:

- author's surname and initials, or name of organisation, or title
- year website was written or last revised
- title of webpage
- date website was accessed
- website address.

For example: Wheeling Jesuit University/Center for Educational Technologies (2013), *NASA Physics Online Course: Forces and Motion*, accessed 16 June 2015, from http://nasaphysics.cet.edu/forces-and-motion.html.

19.3 Review

SUMMARY

- The discussion is the part of the investigation where the evaluation and explanation of the investigation methods and results takes place. It is the interpretation of what the results mean.
- In the discussion, the findings of the investigation need to be analysed and interpreted.
  - State whether a pattern, trend or relationship was observed between the independent and dependent variables. Describe what kind of pattern it was and specify under what conditions it was observed.
  - Were there discrepancies, deviations or anomalies in the data? If so, these should be acknowledged and explained.
  - Identify any limitations in the data collected. Perhaps a larger sample or further variations in the independent variable would lead to a stronger conclusion.
- It is important to discuss the limitations of the investigation method. Evaluate the method and identify any issues that could have affected the validity, accuracy, precision or reliability of the data. Sources of errors and uncertainty must also be stated in the discussion.
- When discussing the results, indicate the range of the data obtained from replicates. Explain how the sample size was selected. Larger samples are usually more reliable, but time and resources are likely to have been scarce. Discuss whether the results of the investigation have been limited by the sample size.
- To make the investigation more meaningful, it should be explained within the right context, meaning the related physics ideas, concepts, theories and models. Within this context, explain the basis for the hypothesis.
- Indicate whether the hypothesis was supported or refuted and on what evidence this is based (that is, the results). Do not provide irrelevant information. Only refer to the specifics of the hypothesis and the research question and do not make generalisations.

KEY QUESTIONS

1. What relationship between the variables is indicated by a sloping linear graph?
2. What relationship exists if one variable decreases as the other increases?
3. What relationship exists if both variables increase or both decrease at the same rate?
4. What might cause a sample size to be limited in an investigation?
5. Consider this investigation hypothesis: An increase in the current passing through a single resistor in an electric circuit will cause an increase in the voltage drop across the resistor.
   Improve this response to the hypothesis:
   When the current was 0.03 A, the voltage was 0.93 V and when the current was 0.05 A, the voltage was 1.81 V.

Chapter review

19

KEY TERMS

controlled variable
dependent variable
independent variable
mean
median
mode
outlier
personal protective equipment (PPE)
qualitative variable
quantitative variable
random error
raw data
reliability
significant figures
systematic error
uncertainty
validity
variable

1 What is a hypothesis and what form does it take?

2 Consider the hypothesis provided below. What are the dependent, independent and controlled variables?
Hypothesis: Releasing an arrow in archery at an angle greater or smaller than 45 degrees will result in a shorter flight displacement (range).

3 What is the dependent variable in each hypothesis?
- **a** If you push an object with a fixed mass (e.g. shot put) with a larger force, then the acceleration of that object will be greater.
- **b** The vertical acceleration of a falling object is constant.
- **c** A springboard diver rotates faster when in a tucked position than when in a stretched (layout) position.

4 List these types of hazard controls from the most effective to the least effective.
substitution, personal protective equipment, engineering controls, administrative controls, elimination, isolation

5 The speed of a toy car rolling down an inclined plane was measured 6 times. The measurements obtained (in cm s^{-1}) were 7.0, 6.5, 6.8, 7.2, 6.5, 6.5.
What is the uncertainty of the average of these values?

6 Which of the statistical measurements of mean, mode and median is most affected by an outlier?

7 What relationship between variables is indicated by a curved trend line?

8 If you hypothesise that impact force is directly proportional to drop height, what would you expect a graph of the data to look like?

9 What is meant by the 'limitations' of the investigation method?

10 What is 'bias' in an investigation?

APPENDIX A SI units

THE STANDARD UNITS OF MEASUREMENT

The accurate and easy measurement of quantities is essential in both everyday life and for scientific investigation. Over the centuries, many different systems of measuring physical quantities have been developed. For example, length can be measured in chains, fathoms, furlongs, yards, feet, rods and microns. Some units were based on parts of the body. The cubit was defined as the distance from the elbow to the fingertip, and so the amount of cloth that you obtained from a tailor depended on the physical size of the person selling it to you.

The metric system was established by the French Academy of Science at the time of the French Revolution (1789–1815) and is now used in most countries. This system includes units such as the metre, litre and kilogram. Countries of the British Empire adopted the British Imperial system of the mile, gallon and pound. These two systems developed independently and their dual existence created problems in areas such as trade and scientific research. In 1960, an international committee set standard units for fundamental physical quantities. This system was an adaptation of the metric system and is known as the *Système Internationale d'Unités* (International System of Units) or SI system of units.

Fundamental quantity	SI unit	SI unit symbol
mass	kilogram	kg
length	metre	m
time	second	s
electric current	ampere	A
temperature	kelvin	K
luminous intensity	candela	cd
amount of substance	mole	mol

TABLE A.1 The SI units identify the seven fundamental quantities whose basic value is defined to a high degree of accuracy.

PHYSICSFILE

Metric system

The metric system was originally developed in France and is known as the *Système Internationale* (SI). It was adopted in France in 1840 as the official system of units, although it had been developing in that country since 1545. It has remained in use ever since and has gradually been adopted by most other countries. It has been modified a little over the years and now, in Australia, we use SI units that have been standardised by the International Standards Organisation (ISO) since the 1960s. Some countries such as France, Italy and Spain use an earlier form of the metric system that is slightly different. The USA still measures almost everything in the old imperial units such as pounds for mass and feet for distance but, even there, scientists use the SI system of units. There are two major advantages of using the metric system. It is easier to use than other systems in that derived units are straightforward and various sizes of units are created using multiples of ten. The other very big advantage is the international nature of the standards and units. All units are standardised, making comparisons straightforward.

Mass

The kilogram was originally defined as the mass of 1 L of water at 4°C. This is still approximately correct, but a far more precise definition is now used. Since 1897 the measurement standard for the kilogram has been a cylindrical block of platinum–iridium alloy kept at the International Bureau of Weights and Measures in France. Australia has a copy of this standard mass at the CSIRO Division of Applied Physics in Sydney. At times it is returned to France to ensure that the mass remains accurate.

Length

The metre was originally defined in 1792 as one ten-millionth of the distance from the equator to the North Pole (approximately 10 000 km). This definition has changed a number of times since. In 1983, to give a more accurate value, the metre was redefined as the distance that light in a vacuum travels in $\frac{1}{299792458}$ second. This standard can be reproduced all over the world, as light travels at a constant speed in a vacuum.

Time

Up to 1967, time had always been based on the apparent motion of the heavens. The second was once defined in terms of the motion of the Sun. Until 1960, one second was defined as $\frac{1}{60}$ of $\frac{1}{60}$ of $\frac{1}{24}$ of an average day in 1900. This reflected the rate of the Earth's rotation on its axis; however, its rotation is not quite uniform. In 1967, a more accurate definition was adopted—one not based on the motion of the Earth. One second is now defined as the time required for a caesium-133 atom to undergo 9 162 631 770 vibrations. These vibrations are stimulated by an electric current and are extremely stable, allowing this standard to be reproduced all over the world.

DERIVED UNITS

As well as the seven fundamental quantities, a wide variety of other physical quantities can be measured. You may have encountered some of these, such as frequency, velocity, energy and density, already. A derived quantity is defined in terms of the fundamental quantities. For example, the SI unit for area is square metres (m^2).

Quantity	SI unit	SI unit symbol	Equivalent unit
velocity	metres per second	$m\ s^{-1}$	—
acceleration	metres per second per second	$m\ s^{-2}$	—
frequency	hertz	Hz	s^{-1}
force	newton	N	$kg\ m\ s^{-2}$
energy/work	joule	J	$kg\ m^2\ s^{-2}$

TABLE A.2 Some derived SI quantities and their units.

APPENDIX B Understanding measurement

MEASUREMENT AND UNITS

In every area of physics we have attempted to quantify the phenomena we study. In practical demonstrations and investigations we generally make measurements and process those measurements in order to come to some conclusions. Scientists have a number of conventional ways of interpreting and analysing data from their investigations. There are also conventional ways of writing numerical measurements and their units.

Correct use of unit symbols

The correct use of unit symbols removes ambiguity, as symbols are recognised internationally. The symbols for units are not abbreviations and should not be followed by a full stop unless they are at the end of a sentence.

Upper-case letters are not used for the names of any physical quantities of units. For example, we write newton for the unit of force, while we write Newton if referring to someone with that name. Upper-case letters are only used for the *symbols* of the units that are named after people. For example, the unit of energy is joule and the symbol is J. The joule was named after James Joule who was famous for studies into energy conversions. The exception to this rule is ‘L’ for litre. We do this because a lower-case ‘l’ looks like the numeral ‘1’. The unit of distance is metre and the symbol is m. The metre is not named after a person.

The product of a number of units is shown by separating the symbol for each unit with a dot or a space. Most teachers prefer a space but a dot is perfectly correct. The division or ratio of two or more units can be shown in fraction form, using a slash, or using negative indices. Most teachers prefer negative indices. Prefixes should not be separated by a space.

Preferred	Correct also	Wrong
m s^{-2}	m.s^{-2} m/s^{2}	ms^{-2}
kW h	kW.h	kWh k Wh
kg m^{-3}	kg.m^{-3} kg/m^{3}	kgm^{-3}
μm		μ m
N m	N.m	Nm

TABLE B.1 Some examples of the use of symbols for derived units.

Units named after people can take the plural form by adding an ‘s’ when used with numbers greater than one. Never do this with the unit symbols. It is acceptable to say ‘two newtons’ but wrong to write 2 Ns. It is also acceptable to say ‘two newton’.

Numbers and symbols should not be mixed with words for units and numbers. For example, twenty metres and 20 m are correct while 20 metres and twenty m are incorrect.

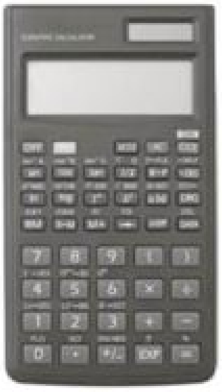


FIGURE B.1 A scientific calculator.

Scientific notation

To overcome confusion or ambiguity, measurements are often written in scientific notation. Quantities are written as a number between one and ten and then multiplied by an appropriate power of ten. Note that 'scientific notation', 'standard notation' and 'standard form' all have the same meaning.

Examples of some measurements written in scientific notation are:

$0.054\text{ m} = 5.4 \times 10^{-2}\text{ m}$

$245.7\text{ J} = 2.457 \times 10^{2}\text{ J}$

$2080\text{ N} = 2.080 \times 10^{3}\text{ N}$ or $2.08 \times 10^{3}\text{ N}$

You should be routinely using scientific notation to express numbers. This also involves learning to use your calculator intelligently. Scientific and graphics calculators can be put into a mode whereby all numbers are displayed in scientific notation. It is useful when doing calculations to use this mode rather than frequently attempting to convert to scientific notation by counting digits on the calculator display. It is quite acceptable to write all numbers in scientific notation, although most people prefer not to use scientific notation when writing numbers between 0.1 and 1000.

An important reason for using scientific notation is that it removes ambiguity about the precision of some measurements. For example, a measurement recorded as 240 m could be a measurement to the nearest metre; that is, somewhere between 239.5 m and 240.5 m. It could also be a measurement to the nearest ten metres, that is, somewhere between 235 m and 245 m. Writing the measurement as 240 m does not indicate either case. If the measurement was taken to the nearest metre, it would be written in scientific notation as 2.40×10^{2} m. If it was taken to the nearest ten metres only, it would be written as 2.4×10^{2} m.

PREFIXES AND CONVERSION FACTORS

Conversion factors should be used carefully. You should be familiar with the prefixes and conversion factors in Table B.2. The most common mistake made with conversion factors is multiplying rather than dividing. Some simple strategies can save you this problem. Note that the table gives all conversions as a multiplying factor.

Multiplying factor		Prefix	Symbol
1 000 000 000 000	10^{12}	tera	T
1 000 000 000	10^{9}	giga	G
1 000 000	10^{6}	mega	M
1 000	10^{3}	kilo	k
0.01	10^{-2}	centi	c
0.001	10^{-3}	milli	m
0.000 001	10^{-6}	micro	µ
0.000 000 001	10^{-9}	nano	n
0.000 000 000 001	10^{-12}	pico	p

Do not put spaces between prefixes and unit symbols. It is important to give the symbol the correct case (upper or lower case). There is a big difference between 1 mm and 1 Mm.

TABLE B.2 Prefixes and conversion factors.

There is no space between prefixes and unit symbols. For example, one-thousandth of an ampere is given the symbol mA. Writing it as m A is incorrect. The space would mean that the symbol is for a derived unit—a metre ampere.

Worked example B1

The diameter of a cylindrical piece of copper rod was measured at 24.8 mm with a vernier caliper. Its length was measured at 35 cm with a tape measure.

a	Find the area of cross-section in m^2.
b	Find the volume of the copper rod in m^3.

SOLUTION

a The area of cross-section is πr^2. The radius is calculated by dividing the diameter by two. Hence the radius is 12.4 mm. To calculate the area in m^2, first halve the diameter and convert it to metres. The radius is $\frac{24.8}{2} = 12.4 \text{ mm} = 12.4 \times 10^{-3}$ m. The radius is not written in scientific notation. This is not necessary. All you need to do is multiply by the appropriate factor. The conversion factor for mm to m is 10^{-3}. Just multiply by the conversion factor and don't bother to rewrite the result in scientific notation. This is because it is only going to be used in a calculation and is not a final result.

The area of cross-section is $\pi r^2 = \pi(12.4 \times 10^{-3})^2 = 4.8 \times 10^{-4}$ m^2.

b The volume is $\pi r^2 h$, where h is the length of the cylinder.
The length is 35 cm = 35×10^{-2} m.
Hence the volume is $\pi(12.4 \times 10^{-3})^2(35 \times 10^{-2}) = 1.7 \times 10^{-4}$ m^2.

Worked example B2

a	A car is traveling at 110 km h^{-1}. How fast is this in m s^{-1}?
b	Convert 35 miles per hour to metres per second. A mile is approximately 1600 m.

SOLUTION

a 110 km h^{-1} is 110×10^3 metres per 3600 s.

$$\frac{110 \times 10^3}{3600} = 30.6$$

Hence 110 km h^{-1} = 30.6 m s^{-1}.

b 35 miles per hour is 35×1600 metres per 3600 s.

$$\frac{35 \times 1600}{3600} = 15.6$$

Hence 35 mph = 16 m s^{-1}.

DATA

Physicists and physics students collect, analyse and interpret experimental data. In fact, you will do this when you conduct your Practical Investigation in Unit 2 (Area of Study 3). Working with data requires a good understanding of the meaning and limitations of measurement.

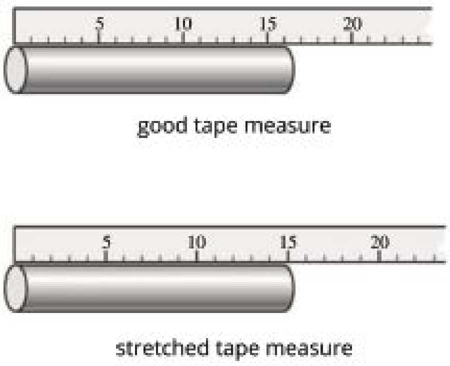


FIGURE B.2 The diagram shows that a correctly manufactured tape measure correctly measures the cylinder to be 16 cm long while the stretched tape measure gives a wrong measurement of 15 cm. The stretched tape measure is inaccurate.

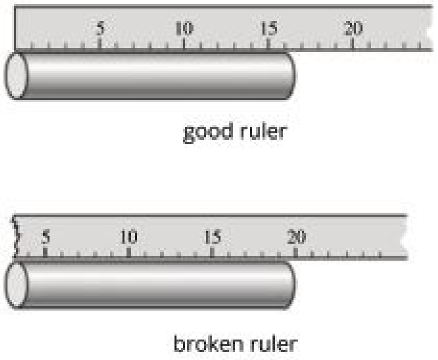


FIGURE B.3 The diagram shows that an undamaged ruler correctly measures the cylinder to be 16 cm long while the broken ruler gives a wrong measurement of 19 cm. The broken ruler is inaccurate but equally as precise as the unbroken ruler.

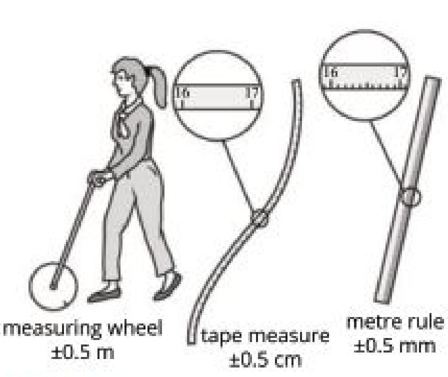


FIGURE B.4 The measuring wheel has low precision and only measures to the nearest metre. It has an uncertainty of 0.5 m. The tape measure has more precision and has an uncertainty of 0.5 cm or 0.005 m. The metre rule has an uncertainty of 0.5 mm or 0.0005 m.

Accuracy and precision

Two very important aspects of any measurement are accuracy and precision. Accuracy and precision are not the same thing. The distinction between the two ideas is only hard to grasp because the two words are defined in a similar way in the dictionary. We often hear the words used together and in general conversation they tend to be used interchangeably.

Instruments are said to be *accurate* if they truly reflect the quantity being measured. For example, if a tape measure is correctly manufactured it can be used to measure lengths accurately to the nearest centimetre.

Imagine that the tape measure is accidentally stretched during the manufacturing process, as shown in Figure B.2. It would still be used to measure length to the nearest centimetre but all measurements would be wrong. It would be inaccurate.

Suppose an accurate ruler had 3 cm snapped off the end, as shown in Figure B.3. It would now give readings all too large by 3 cm if no allowance were made for the missing piece. This ruler measure would be inaccurate.

In these two examples, the tape measure or ruler is used to measure to the nearest centimetre but is inaccurate. Inaccurate means just plain wrong. Instruments are said to be *precise* if they can differentiate between slightly different quantities. Precision refers to the fineness of the scale being used.

Consider the metre rule, the tape measure and the measuring wheel used to mark out sports fields. All three measure distance. All three can be accurate. The metre rule is more *precise* because it measures to the nearest millimetre, the tape measure has less precision due to measuring only to the nearest centimetre, while the wheel measures only to the nearest metre (Figure B.4).

The tape measure is a more precise instrument than the measuring wheel. Suppose two distances of 2673 mm and 2691 mm are being measured with these two instruments. Each distance would be measured as 3 m, to the nearest metre, by the wheel. They would be measured differently as 2.67 m and 2.69 m, to the nearest centimetre, by the tape measure. The tape measure is more precise because it has a finer scale. You might also say that it has greater resolution. The measuring wheel has such low precision that it can't be used to measure which of the two distances is greater or smaller. Measuring instruments with less precision give measurements that are less certain. The uncertainty in the measurement is due to a coarser scale. The measuring wheel gives less certain measurements than the tape measure even though both instruments may be equally accurate.

All measurements have some amount of *uncertainty*, due to the precision of the instrument which does the measuring. (Note that in Chapter 19 the uncertainty due to collecting a range of data was analysed. This section deals with uncertainty due to precision.) The uncertainty is generally one half of the finest scale division on the measuring instrument. The measuring wheel has an uncertainty of 0.5 m. The metre rule has an uncertainty of 0.5 mm. The tape measure has an uncertainty of 0.5 cm. An electronic balance set to measure grams to two decimal places has an uncertainty of 0.005 g.

Sometimes this uncertainty is referred to as error. It is not error, in that it is not a mistake or something wrong. All measuring instruments have limited precision and, in general, the uncertainty is half of the smallest scale division on the instrument.

The uncertainty is, indeed, the measure of the precision of an instrument. It is not related to accuracy. A micrometer screw gauge, which measures length to the nearest one-hundredth of a millimetre and hence is very precise, may not be accurate. Usually they are, but if one has been badly manufactured or bent by being over-tightened repeatedly it most likely will be inaccurate. But its precision will still be ±0.000005 m, or half of one-hundredth of a millimetre.

The uncertainty gives the range in which a measurement falls. If you measured the length of a stick with a metre rule then you would get a measurement 'plus or minus' half a millimetre.

Any stick between 127.5 mm and 128.5 mm long would be measured as 128 mm to the nearest millimetre (refer to Figure B.5). We would record this as 128 ± 0.5 mm.

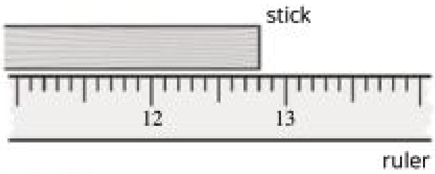


FIGURE B.5 A stick anywhere between 127.5 mm and 128.5 mm would be recorded as having a length of 128 mm if measured by a metre rule with a scale division of 1 mm. Conversely, a measurement recorded as 128 mm could be of an object of length anywhere between 127.5 mm and 128.5 mm.

When using an analogue scale, you might think that you can 'judge by eye' fractions of a scale division and hence get greater precision than half a scale division. You should be able to judge to the nearest half a scale division. You might think you can judge to the nearest tenth of a division. You can't. Research shows that despite the fact that people try to judge the spaces between scale divisions to better than half a division, as soon as this is done, inconsistent measurements are obtained. That is, different people get different measurements of the same thing.

The best judgement you can definitely claim is one half of a scale division. The uncertainty we will still assume, however, is a full half-scale division. Hence, you might measure another stick, one that has a length somewhere between 154 mm and 155 mm, as 154.5 ± 0.5 mm.

Of course, you don't have the option of adding an extra decimal place containing a 0 or a 5 if you are using a digital instrument.

The uncertainty can be recorded as the *absolute* uncertainty as we have done above. The absolute uncertainty is the actual uncertainty in the measurement. In this case it is 0.5 mm. Alternatively, it is often useful to write the uncertainty as a *percentage*: 0.5 mm is 0.32% of 154.5. Hence, the above length would be recorded as 154.5 mm ± 0.32%.

Percentage uncertainty is also called *relative* uncertainty. It is the size of the uncertainty relative to the size of the measured quantity.

PHYSICSFILE

Many people use the term 'error' to refer to uncertainty and many other things. The problem with referring to uncertainty as error is that it is not actually error. Things that are a normal consequence of the limitations of measuring instruments must happen, and are not mistakes. If they are not mistakes or 'something gone wrong' then it makes no sense to call them errors.

Errors are the factors that limit the accuracy of your results. For example, if you perform a calorimetry experiment and do not use a good enough insulator, you will get inaccurate results due to heat losses to the environment. This will contribute to the error in your measurement. Suppose you measured the refraction of light in glass but did not place the protractor in the correct place when measuring angles. This would also cause error.

Many different things can contribute to experimental error. Some are unavoidable. Some are factors in the design of experiments. Good experimental design seeks to eliminate or at least minimise potential sources of error.

Never quote 'human error' as a source of error. Your data should be examined carefully and mistakes eliminated or at least ignored. So-called human errors, or lack of care, have no place in your experimental work. If you make mistakes then you should repeat the measurements.

Estimating the uncertainty in a result

An experiment or a measurement exercise is not complete until the uncertainties have been analysed. Chapter 19 explained how uncertainties were treated when a range of data had been collected and then averaged out. It is also important to explain how uncertainty due to the precision of instruments affects results.

The following three processes are used for estimating uncertainty in calculations due to the precision of instruments. They are demonstrated in Worked example B3.

- When adding or subtracting data, add the absolute uncertainties.
- When multiplying or dividing data, add the percentage uncertainties.
- When raising data to power n, multiply the percentage uncertainty by n.

In Worked example B3, the analysis of uncertainty reveals the *precision* of an experimental result.

Worked example B3

In Area of Study 1 in Unit 1, you might have measured the specific heat capacity of a metal. You could have calculated your result using:

$$c_{\text{metal}} = \frac{c_{\text{water}} m_{\text{water}} \Delta T_{\text{water}}}{m_{\text{metal}} \Delta T_{\text{metal}}}$$

Suppose you had the following data included in your table.

Quantity		Absolute uncertainty	% uncertainty
c_{water}	4180 J kg^{-1} K^{-1}	5 J kg^{-1} K^{-1}	0.120
m_{water}	72.5×10^{-3} kg	0.05×10^{-3} kg	0.069
ΔT_{water}	5°C	1°C*	20
m_{metal}	87.3×10^{-3} kg	0.05×10^{-3} kg	0.057
ΔT_{metal}	72°C	1°C*	1.389

*Note that the ΔT values have an absolute uncertainty of 1°C because they are calculated by subtracting one temperature measurement from another.

You would calculate as follows:

$$\begin{aligned} c_{\text{metal}} &= 241 \text{ J kg}^{-1} \text{ K}^{-1} \\ \text{Uncertainty (\%)} &= 0.120 + 0.069 + 20 + 0.057 + 1.389 \\ &= 21.6\% \end{aligned}$$

Hence, you would obtain the following result:

$$\begin{aligned} c_{\text{metal}} &= 241 \text{ J kg}^{-1} \text{ K}^{-1} \pm 21.6\% \\ c_{\text{metal}} &= 241 \pm 52 \text{ J kg}^{-1} \text{ K}^{-1} \end{aligned}$$

Once you have done all of this you can consider the relative success of your measurement exercise.

Your result is:

$$189 \text{ J kg}^{-1} \text{ K}^{-1} \le c_{\text{metal}} \le 293 \text{ J kg}^{-1} \text{ K}^{-1}$$

If measurements by other people, such as the constants published in data books, fall within this range then you can conclude that your experiment is consistent with established values. That is, within the precision of your technique, there are probably no significant errors although the final measurement is rather imprecise in this case. We might say that it is accurate within the limitations of the equipment.

PHYSICSFILE

In some classes, students are instructed to quote all results to two decimal places or to three significant figures. You should be able to see from Worked example B3 that these rules are not absolutely correct when applied to real data. For ordinary calculations in assignments, tests and examinations, you might just give your answers to three figures.

If a calculation is done in several stages then you should not round off any intermediate results. This will add rounding error to your calculations. Use the memory on your calculator so that there is no rounding until the end of your calculation.

You are also now in a position to refine the experiment by reducing the larger uncertainties. In this case, the largest uncertainty was in the temperature change for the water. Hence, it would not be very helpful to measure the masses to greater precision because the limit to precision in this activity would be the temperature differences. Getting greater precision in the temperature changes would be a useful refinement.

You could consider ways of getting larger temperature changes in the water and hence obtain a smaller percentage uncertainty in the temperature change. Alternatively, you might consider ways of measuring the temperatures to greater precision.

If your measurement range does not include the result you expect, you should think about the origin of the errors. In other words, if you are sure that c_{metal} is less than 189 J kg^{-1} K^{-1} or more than 293 J kg^{-1} K^{-1} then there must be some error in your experimental technique or more uncertainty than you realised.

When reviewing an experiment or a measurement exercise, it is a good idea to consider both errors *and* uncertainties.

Significant figures

The number of significant figures in a measurement is simply the number of digits used when the number is written in scientific notation. (Note: Significant figures were explained in Chapter 19.) Once you have done a calculation, your calculator usually has eight or ten digits in the display but most of them are meaningless. You must round off your answer appropriately.

Consider the result of the experiment described in Worked example B3. It would make no sense to quote the result to two decimal places (or five significant figures) when clearly the precision of the experiment gives less than three significant figures.

Calculated results never have more significant figures than the original data and might have fewer than the original data. If you are not doing a full analysis of the uncertainties, it is customary to give your answers to the same number of significant figures as the least precise piece of data. For example, in Worked example B3, the least precise data is the change in temperature of the water with only a single digit. The value for the specific heat might then be quoted simply as 2×10^2 J kg^{-1} K^{-1}, but doing the full calculation of the uncertainty in the result is much more informative.

GRAPHICAL ANALYSIS OF DATA

A major problem with doing a calculation from just one set of measurements is that a single incorrect measurement can significantly affect the result. Scientists like to take a large amount of data and observe the trends in that data. This gives more precise measurements and allows scientists to recognise and eliminate problematic data.

Physicists commonly use graphical techniques to analyse a set of data. In this section, the basic techniques that they use will be outlined and a general method for using a set of data that fits a known mathematical relationship will be developed.

Linear relationships

Some relationships studied in physics are linear, that is a straight line, while others are not. It is possible to manipulate non-linear data so that a linear graph reveals a measurement. Linear relationships and their graphs are fully specified with just two numbers: gradient, m, and vertical axis intercept, c. In general, linear relationships are written:

$$y = mx + c$$

The gradient, m, can be calculated from the coordinates of two points on the line:

$$m = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1}$$

where (x_1, y_1) and (x_2, y_2) are any two points on the line. Don't forget that m and c have units. Omitting these is a common error.

PHYSICSFILE

Graphs

When analysing data from a linear relationship, it is first necessary to obtain a graph of the data and an equation for the line that best fits the data. This line of best fit is often called the regression line. The entire process can be done on paper but most people will use a computer spreadsheet, the capabilities of a scientific or a graphics calculator, or some other computer-based process. In what follows, it is not assumed that you are using any particular technology.

If you are plotting your graph manually on paper then proceed as follows:

1. Plot each data point on clearly labelled, unbroken axes.
2. Identify and label but otherwise ignore any suspect data points.
3. Draw, by eye, the 'line of best fit' for the points. The points should be evenly scattered either side of the line.
4. Locate the vertical axis intercept and record its value as '*c*'.
5. Choose two points on the line of best fit to calculate the gradient. Do not use two of the original data points as this will not give you the gradient of the line of best fit.
6. Write $y = mx + c$, replacing x and y with appropriate symbols, and use this equation for any further analysis.

If you are using a computer or a graphics calculator then proceed as follows:

1. Plot each data point on clearly labelled, unbroken axes.
2. Identify suspect data points and create another data table without the suspect data.
3. Plot a new graph without the suspect data. Keep both graphs as you don't actually discard the suspect data but do eliminate it from the analysis.
4. Plot the line of best fit—the regression line. The manner in which you do this depends on the model of calculator or the software being used.
5. Compute the equation of the line of best fit that will give you values for m and c.
6. Write $y = mx + c$, replacing x and y with appropriate symbols, and use this equation for any further analysis.

Worked example B4

Some students used a computer with an ultrasonic detector to obtain the speed–time data for a falling tennis ball. They wished to measure the acceleration of the ball as it fell. They assumed that the acceleration was nearly constant and that the relevant relationship was $v = u + at$, where v is the speed of the ball at any given time, u was the speed when the measurements began, a is the acceleration of the ball and t is the time since the measurement began.

Their computer returned the following data:

Time (s)	Speed (m s^{-1})
0.0	1.25
0.1	2.30
0.2	3.15
0.3	4.10
0.4	5.25
0.5	6.10
0.6	6.95

Find their experimental value for acceleration.

Solution

The data is assumed linear, with the relationship $v = u + at$, which can be thought of as being $v = at + u$, which makes it clear that putting v on the vertical axis and t on the horizontal axis gives a linear graph with gradient a and vertical intercept u. A graph of the data is shown in Figure B.6.

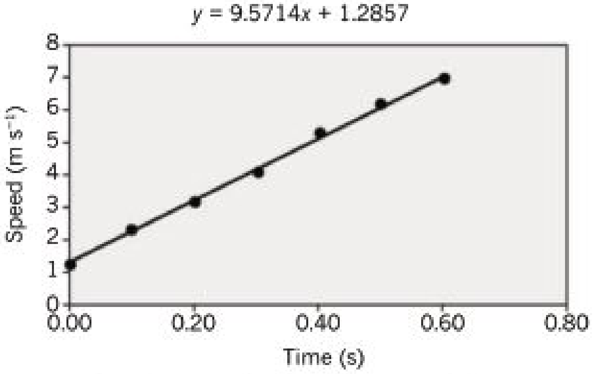


FIGURE B.6 Speed–time profile for a falling tennis ball.

This graph of the data was created on a computer spreadsheet. The line of best fit was created mathematically and plotted. The computer calculated the equation of the line. Graphics calculators can also do this.

A scientific calculator or graphics calculator or spreadsheet gives the regression line as $y = 9.5714x + 1.2857$. If this is rearranged and the constants are suitably rounded, the equation is $v = 1.3 + 9.6t$. This indicates that the ball was moving at 1.3 m s^{-1} at the commencement of data collection and the ball was accelerating at 9.6 m s^{-2}.

Manipulating non-linear data

Suppose you were examining the relationship between two quantities B and d and had good reason to believe that the relationship between them is

$$B = \frac{k}{d}$$

where k is some constant value. Clearly, this relationship is non-linear and a graph of B against d will not be a straight line. By thinking about the relationship it can be seen that in 'linear form':

$$B \;=\; k\;\frac{1}{d}$$
$$\uparrow \qquad \uparrow\;\uparrow$$
$$y \;=\; m\;x + c$$

A graph of B (on the vertical axis) against $\frac{1}{d}$ (on the horizontal axis) will be linear. The gradient of the line will be k and the vertical intercept, c, will be zero. The line of best fit would be expected to go through the origin because, in this case, there is no constant added and so c is zero.

In the above example, a graph of the raw data would just show that B is larger as d is smaller. It would be impossible to determine the mathematical relationship just by looking at a graph of the raw data.

A graph of raw data will not give the mathematical relationship between the variables but can give some clues. The shape of the graph of raw data may suggest a possible relationship. Several relationships may be tried and then the best is chosen. Once this is done, it is not proof of the relationship but, possibly, strong evidence.

When an experiment involves a non-linear relationship, the following procedure is followed:

1 Plot a graph of the original raw data.
2 Choose a possible relationship based on the shape of the initial graph and your knowledge of various mathematical and graphical forms.
3 Work out how the data must be manipulated to give a linear graph.
4 Create a new data table.

Then follow the steps given in the Physics file on page 490. It may be necessary to try several mathematical forms to find one that seems to fit the data.

Worked example B5

Some students were investigating the relationship between current and resistance for a new solid-state electronic device. They obtained the data shown in the table.

According to the theory they had researched on relevant Internet sites, the students believed that the relationship between I and R is $R = dI^3 + g$, where d and g are constants.

By appropriate manipulation and graphical techniques, find their experimental values for d and g. The following steps should be used:

a Plot a graph of the raw data.
b Work out what you would have to graph to get a straight line.
c Make a new table of the manipulated data.
d Plot the graph of manipulated data.
e Find the equation relating I and R.

Current, I (A)	Resistance, R (Ω)
1.5	22
1.7	39
2.2	46
2.6	70
3.1	110
3.4	145
3.9	212
4.2	236

Solution

a Figure B.7 shows the graph obtained using a spreadsheet.
It might be argued that the second piece of data is suspect. The rest of this solution supposes the students chose to ignore this piece of data.

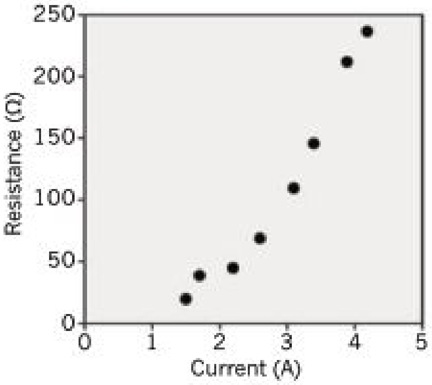


FIGURE B.7 Current–resistance characteristic.

b You can see what to graph if you think of the equation like this:

$$\begin{array}{cccccc} R & = & d & I^3 & + & g \\ \uparrow & & \uparrow & \uparrow & & \uparrow \\ y & = & m & x & + & c \end{array}$$

A graph of R on the vertical axis and I^3 on the horizontal axis would have a gradient equal to d and a vertical axis intercept equal to g.

c The data is manipulated by finding the cube of each of the values for current.

Current cubed, I^3 (A^3)	Resistance, R (Ω)
3.38	22
10.65	46
17.58	70
29.79	110
39.30	145
59.32	212
74.09	236

d The graph in Figure B.8 was obtained from the spreadsheet.

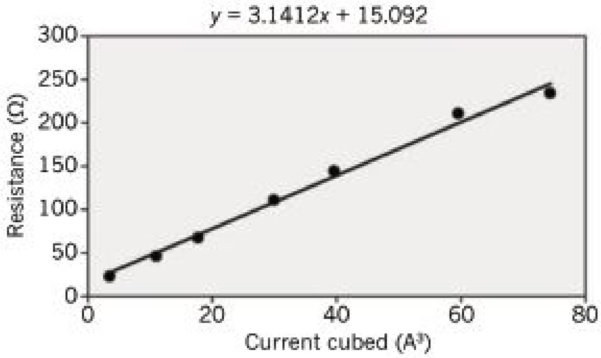


FIGURE B.8 Current–resistance characteristic (manipulated data).

e The regression line has the equation $y = 3.1x + 15.1$, so the equation relating I and R is $R = 3.1I^3 + 15.1$. Hence, the value of d is 3.1 Ω A^{-3} and the value of g is 15.1 Ω.

Answers

Chapter 1 Heating Processes

1.1 Heat and temperature

WE 1.1.1 $\Delta U = 2770$ J
1 C **2** **a** C and D **b** B **3** C and D
4 The temperature of the gas is just above absolute zero so the particles have very little energy.
5 **a** 303 K **b** 102°C
6 The average kinetic energy of the hydrogen particles in tank B is greater than the average kinetic energy of the hydrogen particles in tank A.
7 absolute zero, 10 K, −180°C, 100 K, freezing point of water
8 −70 kJ **9** 225 J
10 $W = -550$ J so the scientist does 550 J of work on the sodium.

1.2 Specific heat capacity

WE 1.2.1 6.3 MJ **WE 1.2.2** 2:1
1 Water requires the most energy to achieve this result.
2 The aluminium has the most thermal energy.
3 2100 J **4** 25 200 J or 25.2 kJ **5** 2x J
6 The temperature of the aluminium is 4.67 times that of the water.
7 B **8** 10.0 kJ **9** 30°C **10** 1 kg

1.3 Latent heat

WE 1.3.1 $Q = 1.38 \times 10^2$ kJ
WE 1.3.2 $L_{vapour} = 22.5 \times 10^5$ J kg^{-1}
1 The mercury is changing state from solid to liquid—it is melting.
2 −39°C **3** 357°C **4** $L = 1.26 \times 10^4$
5 $L = 2.85 \times 10^5$ J kg^{-1} **6** $Q = 2.25 \times 10^5$ J
7 34 kJ **8** B **9** D

1.4 Conduction

1 The mass of the particles is relatively large and the vibrational velocities are fairly low.
2 Metals conduct heat by free moving electrons as well as by molecular collisions. Wood does not have any free moving electrons, so it is a poor conductor of heat.
3 Thickness, surface area, nature of the material and the temperature difference between it and the second material.
4 Copper is a better conductor of heat than stainless steel.
5 C
6 A lot of air is trapped in the down. As air is a poor conductor of heat, the down-filled quilt limits the transfer of heat away from the person.
7 escaping; low; not able
8 Plastic and rubber have low conductivity, so they do not allow the transfer of heat from your hand very easily. Metal has high conductivity, so heat transfers from your hand easily and your hand feels cold.

1.5 Convection

1 liquids and gases
2 upwards
3 Air over certain places, such as roads, heats up and as a result becomes less dense. The less dense air rises to form a column of rising air called a thermal.
4 D
5 It is not possible for solids to pass on heat by convection because solids do not contain the free molecules that are required to establish convection currents.
6 The source of heat, the Sun, is at the top of the water. It takes much longer to heat a liquid when the source is at the top as the convection currents will also remain near the top. The warm water is less dense than the cool water and will not allow convection currents to form throughout the water.

1.6 Radiation

1 **a** partially reflected, partially transmitted, partially absorbed
b absorption
2 higher; shorter; infrared **3** E
4 Conduction and convection require the presence of particles to transfer heat. Heat transfer by radiation can occur in a vacuum as the movement of particles is not required.
5 The person, due to their temperature, emits stronger infrared radiation than their surroundings.
6 **a** the matt black beaker **b** the gloss white beaker
7 dark-coloured metals

Chapter 1 Review

1 A **2** temperature
3 Heat refers to the energy that is transferred between objects, whereas temperature is a measure of the average kinetic energy of the particles within a substance.
4 **a** 278 K **b** −73°C
5 The balls will be at the same temperature. **6** B
7 A change of state. Both have the same kinetic energy as their temperatures are the same; however, the steam has more potential energy due to its change in state.
8 The steam. It occurs because the heat energy is used to increase the potential energy of the particles in the solid instead of increasing their kinetic energy. The energy needed to change from solid to liquid is the latent heat of fusion.
9 The higher energy particles are escaping, leaving behind the lower energy particles. The result is that the average kinetic energy of the remaining particles decreases, thus the temperature drops.
10 $c = 126$ J kg^{-1} K^{-1} **11** $Q = 7.0$ kJ **12** $T = 60$°C
13 Polystyrene is a good insulator, whereas metal is a good conductor of heat.
14 As sweat evaporates, it cools down.
15 The stopper reduces heat loss by convection and conduction.
16 Water is a better conductor of heat than air. A wetsuit provides a layer of insulating material (neoprene) around the body and slows the rate of heat loss from the warm body into the cold water.

Chapter 2 Applying thermodynamic principles

2.1 Heating by radiation

WE 2.1.1 32 200 K **WE 2.1.2** 9.66 μm
WE 2.1.3 The rate of transmission is increased by 16 times.
WE 2.1.4 952 W
1 infrared **2** $T = 3622.5$ K **3** $T = 4140$ K
4 $\lambda_{max} = 322$ **5** D
6 A bald head will absorb radiant energy at 1/3 the rate.
7 $P = 79$ W **8** $T = 119$ W **9** $T = 15$°C

2.2 The enhanced greenhouse effect

1 the Sun
2 visible; longer; infrared
3 Carbon dioxide, like other greenhouse gases, absorbs and emits long-wavelength infrared radiation, and re-radiates it back to the surface of the Earth rather than reflecting it back out to space. Carbon dioxide is present in the largest proportion relative to other greenhouse gases so has a greater impact than other greenhouse gases.
4

Greenhouse gas	Source of greenhouse gas
carbon dioxide	combustion of fossil fuels
methane	agriculture
CFCs	air-conditioners
nitric oxide	artificial fertilisers

5 Thermal energy moves through the Earth's mantle by convection currents that rise to the upper layer of the mantle and fan out before sinking as cooler magma.

6 While some energy is transferred by conduction and radiation in the atmosphere, the major method is by convection. Locally, sea breezes and land breezes move heat energy around. On a global scale, radiant energy from the Sun at the equator heats the air. Cool, dense air moves in, forcing the less dense warm air upwards where it spreads out towards the poles. There it cools and sinks.

7 On a hot day, the surface of a land mass heats up more quickly than the ocean. Hot air rises over the land and cooler, denser air moves in from over the ocean creating a breeze.

2.3 Scientific modelling: The enhanced greenhouse effect

1 A

2 The time of year must be kept constant because there may be seasonal changes in methane concentration.

3 In order to remove bias from data, standard procedures need to be followed. Without this, the data may be invalid.

4 **1** make an observation **2** form a hypothesis
  3 apply/test the hypothesis
  4 assess the reliability of the results
  5 carbon dioxide; warmer; ice ages; interglacial periods

2.4 Issues related to thermodynamics

1 windows **2** to allow maximum exposure to the Sun

3 E **4** conduction, convection and radiation

5 ethanol, LPG, petrol, diesel

6 on average 38%, or just over one-third

Chapter 2 Review

1 As temperature increase, the wavelength of the emitted radiation decreases and the frequency increases.

2 gamma rays (γ-rays)
X-rays ultraviolet (UV) visible light infrared (IR)
microwaves radio and television waves

3 B **4** It will decrease by 16 times.

5 **a** The heater is a very good absorber of radiant thermal energy.
  b The heater is also a good emitter of radiant energy.

6 $\lambda_{max} = 414$ nm

7 The first kettle will emit radiant thermal energy at double the rate of the second kettle.

8 C **9** $\lambda_{max} = 10$ μm **10** $T = 4500$ K

11 e = 0.8 **12** UV

13 The increase in the retention of radiated thermal energy by the Earth's atmosphere due to the increase of greenhouse gases.

14 Building cities and land clearing on a large-scale changes the amount of thermal energy retained or reflected by changing the surface characteristics by increasing the emissivity.

15 A

16 The sun is higher in the sky during the summer months and so is blocked by the eaves. The lower angle of the winter sun means that sunlight can pass under the eaves and through the windows.

17 Deciduous trees lose their leaves in winter, allowing sunlight to strike a window. In summer, deciduous trees are full of leaves, providing shade for the window

18 reduce; convective; reducing; 3°C

19 Running an electric car produces no direct emissions. Recharging from a zero emissions source such as solar power would make an electric car emission free.

Unit 1 Area of Study 1 Review

1 D

2 There has to be thermal contact for equilibrium to be reached. This does not have to be physical contact because heat can be exchanged via radiation from a distance.

3 B

4 Initially the air is a gas at room temperature with relatively high average kinetic energy. Energy flows from the gas to the liquid nitrogen. Gas molecules lose so much kinetic energy that they no longer have enough energy to overcome their attraction with one another, and the gas condenses to a liquid. Gas volume in the balloon is reduced. On removal from the liquid nitrogen, the molecules inside the balloon absorb energy from the surrounding air. Upon heating, the liquid boils and vaporises. Temperature and pressure are restored.

5 B **6** C

7 Heat is absorbed from the environment, increasing the internal energy of the CO_2 molecules and changing the intermolecular bonds.

8 Temperature is related to the average kinetic energy of the particles. On sublimation, the average kinetic energy of the particles is not altered. Potential energy increases as the molecules move further apart.

9 aluminum **10** C

11 Energy flows from an object at a higher temperature to that at a lower temperature until both furnace and copper are at the same temperature. The amount of energy that leaves the furnace is equal to the amount of energy that enters the copper rod.

12 When energy is added to the rod, the atoms move faster, increasing their average kinetic energy. Temperature is related to the average kinetic energy of the particles, and so the temperature increases.

13 The rod will lose heat by radiation and conduction to the steel plate and the nitrogen gas.
(No convection in the steel or the copper solids, but convection currents in the gas could assist with cooling.)

14 B and C

15 When heat is absorbed by a material and no phase change is involved, the heat capacity is the energy in joules to heat 1 kg of material by one degree. For phase changes (latent heat) there is no temperature change—it is merely the energy per kg to cause the phase change.

16 $Q = 47.5$ kJ

17 Water has a very high specific heat capacity relative to the fats and proteins in the ice cream. Ice cream mix is only 70.0% water, hence its specific heat capacity is lower than that of pure water.

18 The heat lost to the brine at 0°C comes from the latent heat of fusion and the water changes phase and becomes ice.

19 $Q = 116.9 = 117$ kJ **20** $T_f = -1.6$°C

21 $Q = 199$ kJ **22** $Q = 262.6$ kJ

23 $Q = 52.8$ kJ **24** mass of nitrogen = 0.114 kg

25 mass of ice = 167 g **26** $SHC_c = 462 \text{ J kg}^{-1} \text{ K}^{-1}$

27 Water has an exceptionally high specific heat capacity, and thus can absorb large amounts of heat without significantly changing temperature. This moderates the temperatures close to the coast.

28 Obviously the amount of steam or water in question would make a difference since the heat is proportional to the mass.
If the masses were equal, the steam will burn more severely because of the additional latent heat that is released when it condenses to water on the person's skin at 100°C.

29 D **30** $T_i = 32.7$°C

31 **a** radio waves, microwaves, infrared, visible light, ultraviolet, X-rays, gamma rays
  b gamma rays, X-rays, ultraviolet, visible light, infrared, microwaves, radio waves

32 Radiation: heat enters the Earth's atmosphere via heat radiated by Sun and is re-radiated out by black-body radiation; some of the radiant heat is trapped creating the greenhouse effect.
Conduction: air is a poor conductor of heat.
Convection: major factor; hot air rises, cools and drops; basis of most weather systems.

33 B

34 **a** Incident: $P = 1370 \text{ W m}^{-2} \times \pi R_E^2$
  b Absorption: $P = 945 \text{ W m}^{-2} \times \pi R_E^2$

35 $P = 5.67 \times 10^{-8} \text{ J K}^{-4} \times T_E^4 \times 4\pi R_E^2$

36 The effect of the atmosphere is to retain some of the reflected heat. This raises the average temperature.

37 **a** Apart from air temperature, it is the heat re-radiated from the ground that affects the night temperatures. Clouds absorb some of the radiated heat and reflect it back to Earth; on clear nights, more energy is lost by radiation.

b Energy lost = 1.20 MJ

38 The indoor heater heats the air at the burner by conduction, and forces the hot air through the vents, where convection will carry it throughout the house. People are heated by conduction as the warm air contacts their skin. The outdoor heater is primarily radiative. People are heated by the radiation from the heater far more than by conduction or convection of air. So the salesman is correct.

39 The plate radiates heat as it glows, some of this heats the pot. The plate also transfers heat to the pot and to the water by conduction. Convection currents set up in the water. The peas are heated by the conduction of heat from the hot water.

40 The greenhouse effect means that Venus re-radiates very much less energy. The CO_2 in the atmosphere absorbs much of the re-radiated energy which traps the heat on the planet, causing a higher temperature.

41 As $P \propto T^4$, doubling the temperature increases the radiant power by a factor of 16.

42 TB = 3300 K

43 A human body radiates energy in the infrared region of the electromagnetic spectrum. The goggles are sensitive to these wavelengths, and so people can be detected by the radiation they emit.

44 **a** λ_{max} = 9.99 μm **b** T = 700 K

45 A black body emits a whole range of frequencies. In this case the peak is in the infrared. There will be emissions at shorter wavelengths that are in the red end of the visible spectrum.

46 Wien's Law: $\lambda_{max}T$ = constant

As the temperature goes up, the peak of the black-body spectrum moves to shorter wavelengths, going from red to yellow to bright white. At lower temperatures the emission of radiation is outside the visible range.

47 D

48 The greenhouse effect simply refers to the role of gases such as CO_2, which absorb some of the energy re-radiated by the Earth, and results in an increased temperature. The enhanced greenhouse effect refers to the increasing levels of CO_2, which are usually attributed to human activity since the Industrial Revolution, in particular the burning of fossil fuels.

49 Yes, data from various remote observation stations (Cape Grim, Mauna Loa), and ice core samples show a steady increase in CO_2.

50 The global climate system is very complex. There are natural variations and cycles, as well as long term trends. There are many factors that influence climate, and it is difficult to attribute a change to only one factor.

Chapter 3 Electrical physics

3.1 Behaviour of charged particles

WE 3.1.1 -6.4×10^{-13} C **WE 3.1.2** 3.0×10^{13} electrons

1 They will attract as they will be oppositely charged.

2 3.1×10^{19} **3** +6.7 C

4 Copper is a good conductor of electricity because its electrons are loosely held to their respective nuclei. This allows electrons to move freely through the material by 'jumping' from one atom to the next. Rubber is a good insulator. The rubber coating is used to insulate copper wiring to prevent charge leaving the circuit.

3.2 Electric current and circuits

WE 3.2.1 4.69×10^{18} electrons

1 A continuous conducting loop (closed circuit) must be created from one terminal of a power supply to the other terminal.

2 cell, light bulb, open switch, resistor and ammeter

3 C **4** **a** 3 A **b** 0.5 A **c** 0.008 A

5 **a** 5 C **b** 300 C **c** 18 000 C

6 **a** 9 A to the right **b** 2 A to the left

7 **a** 3 C **b** 1000 C **c** 1440 C **8** **a** 16 C **b** 4 A

9 **a** 2×10^{19} **b** 0.32 A **10** B

3.3 Energy in electric circuits

WE 3.3.1 43 200 J **WE 3.3.2** 9450 J **WE 3.3.3** 720 W

1 A **2** **a** 138 000 J (or 138 kJ) **b** 2 A

3 **a** **i** 4 V **ii** 4 V **iii** 2 V **b** **i** 10 A **ii** 1 A **iii** 1 A

4 20 V **5** 167 C

6 **a** heat and light **b** 60 W **c** 0.25 A

7 the gravitational potential energy of the water

8 **a** M2 or M3 **b** M1 or M4

3.4 Resistance

WE 3.4.1 24 Ω **WE 3.4.2** I_1 = 0.6 A, V_2 = 12 V

WE 3.4.3 778 Ω **WE 3.4.4** 9 600 mA **WE 3.4.5** 4.0 V

1 **a** A, B, C **b** C, B, A **2** 0.375 A 4.8 V

3 **a** The wire is ohmic. This is because there is a proportional relationship between the voltage and the current, as shown by the linear nature of the I–V graph, which means that the resistance is a constant.

b 3 A **c** 2.5 Ω

4 **a** 0.71 Ω **b** The resistor is ohmic.

5 They are both right. The resistance of the device is different for different voltages. Therefore the device is non-ohmic.

6 72 mA **7** **a** 2 Ω **b** 5 A **8** a 1.6 Ω **b** 0.2 Ω

9 **a** It is non-ohmic, as the I–V relationship is non-linear.

b 0.5 A **c** 15 V

d The resistance of the device at these voltages will be given by $R = V/I$.

i 20 Ω. **ii** 13.3 Ω

Chapter 3 Review

1 1.9×10^{19} electrons **2** 6.7 C **3** A

4 $Q = 3.2 \times 10^{-19}$ C **5** C **6** D

7 3.8×10^{-3} A

8 Conventional current represents the flow of charge around a circuit as if the moving charges were positive, which means the direction is from the positive terminal to the negative terminal. In reality, the moving particles in a metal wire are negatively charged electrons. Electron flow describes the movement of these electrons from the negative terminal to the positive terminal.

9 **a** 160 C **b** 10^{21} electrons **10** a 0.8 C **b** 20 s

11 7.6 J **12** 4 V **13** C **14** 1.39 W

15 8.75 V **16** 8.7 A **17** 0.5 Ω **18** 960 Ω

19 333 Ω **20** 20 V **21** 30 Ω

22 **a** 1 Ω **b** 2 Ω **c** 3 Ω

23 As electrons travel through a piece of copper wire, they constantly bump into copper ions that slow them down. Resistance is a measure of how much energy electrons need to be given to maintain a constant speed through the wire.

24 216 kJ **25** 0.4 Ω **26** 120 W **27** 36 Ω

28 When lamp is cold, R = 9.6 Ω. When lamp is hot, R = 150 Ω.

29 5.0×10^{18} electrons

30 **a** 18 C **b** 54 J **c** The energy is provided by the battery.

Chapter 4 Practical electrical circuits

4.1 Series and parallel circuits

WE 4.1.1 160 Ω

WE 4.1.2 I = 0.011 A, V_1 = 1.1 V, V_2 = 7.6 V, V_3 = 3.6 V

WE 4.1.3 14.3 Ω

WE 4.1.4 $I_{circuit}$ = 0.53 A, I_{30} = 0.33 A, I_{50} = 0.2 A

WE 4.1.5 ΔV_1 = 29.6 V

$\Delta V_{2-4} = 1.48 \times 14.3 = 21.2$ V $\Delta V_{5-6} = 1.48 \times 3.3 = 4.9$ V

$\Delta V_7 = 1.48 \times 30.0 = 44.4$ V $I_1 = I_7 = 1.48$ A $I_2 = 0.42$ A

$I_3 = 0.85$ A $I_4 = 0.21$ A $I_5 = 0.98$ A $I_6 = 0.49$ A

WE 4.1.6 in parallel, P = 0.9 W, in series, P = 0.144 W

The parallel circuit draws 6.25 times as much power as the series circuit.

1 B **2** **a** $I_T = 0.0075$ A or 7.5 mA **b** $V_{100} = 0.75$ V

3 $R_1 = R_2 = 136\ \Omega$

4 **a** $I_T = 0.75$ A **b** $I_{20} = 0.25$ A **c** $I_{20} = 0.5$ A

5 **a** $V_T = 12$ V **b** $I_{60} = 0.2$ A (or 200 mA)

6 $I_1 = I_2 = 0.3$ A $V_1 = 6$ V $V_2 = 4.5$ V
$V_3 = V_4 = 1.5$ V $I_3 = I_4 = 1.5$ A

7 $R = 4.278\ \Omega$ **8** **a** $P = 1.25$ W **b** $P = 20$ W **9** C

4.2 Using electricity

WE 4.2.1 4 V **WE 4.2.2** **a** $R = 20$ kΩ, **b** $I = 0.20$ mA, **c** $V = 1$ V

WE 4.2.3 **a** $R = 9.8$ kΩ, **b** $I = 0.92$ mA, **c** $V = 6.3$ V

WE 4.2.4 **a** 0.6 V, **b** 1680 Ω **WE 4.2.5** 137 Ω

1 **a** C **b** A **2** $V_{2000} = 16$ V

3

Input transducer	Signal-processing component	Output transducer
LDR / microphone / thermistor	diode / potentiometer	LED / light globe / speaker

4 A and B **5** B **6** B **7** A

8 $R_L = 350\ \Omega$

9 **a** Circuit (i): $R_L = 117\ \Omega$ Circuit (ii): $R_L = 150\ \Omega$

b Both circuits will emit the same light—all are operating at the same V and I.

c Circuit (ii) requires only 20 mA, so will run longer.

10 **a** $R_T = 500\ \Omega$ **b** $R_R = 625\ \Omega$ **c** $T = 10$°C

4.3 Electrical safety

WE 4.3.1 \$1.63

1 A **2** D **3** 10 kWh = 3.6×10^7 J

4 This air conditioner would cost 0.75 × 5 × 0.27 = approximately \$1 to run for 5 hours. Therefore, the figure \$10 in the statement is incorrect.

5 The neutral and earth are common.

6 It is much safer to place the fuse in the active circuit because then it cuts off the supply to the circuit.

7 The earth stake ensures that the neutral and earth conductors are at zero potential.

8 The toaster will work normally, but the connection is very unsafe because it will remain live even when switched off.

9 The outer casing of the appliance could become live.

10 $I = 2.4$ mA

Chapter 4 Review

1 A

2 **a** $R_3 = 3.5\ \Omega$ **b** $I_3 = 0.35(3)$ A **c** $V_{\text{parallel pair}} = 1.8$ V

d $I_2 = 0.12$ A **e** $I_1 = 0.24$ A **f** $R_1 = 7.46\ \Omega$

3 **a** Ammeter. **b** $R_{\text{total}} = 8.57\ \Omega$

4 The earth wire is usually connected to the metal casing of an electrical appliance. If the insulation around the wire inside the appliance becomes degraded, the casing of the appliance could become 'live' and dangerous to touch. In this situation, the earth wire provides an alternative low-resistance path to earth, protecting any users of the appliance from electrocution.

5 The circuit will need to have either two pairs of series resistors connected in parallel or two pairs of parallel resistors connected in series.

6 $V_{600} = 3$ V **7** B **8** C

9 **a** $P = 2.4$ W **b** $P = 21.6$ W

c The parallel circuit draws 9 times more power.

10 D **11** Cost = \$3.60 **12** D

13 **a** $R_T = 60\ \Omega$ **b** $I = 2.0$ A

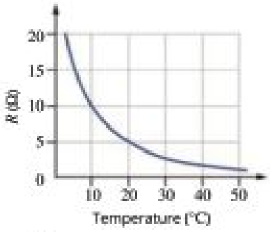
c $I_1 = 1.20$ A, $I_2 = 0.60$ A, $I_3 = 0.20$ A

d $P = 240$ W **e** 240 W

14 **a** 4 V **b** above **c** V_{out} approaches zero

15 **a** A thermistor is a temperature-sensitive resistor whose resistance decreases as its temperature increases. It is usually constructed from Ge, Si or a mixture of various oxides. The resistance of a typical thermistor may range from 10 kΩ at 0°C to 100 Ω at 100°C.

b



not linear

16 $T = 20$°C **17** **a** $V_{\text{total}} = 75$ V **b** $P = 37.5$ W

18 **a** 3 **b** 1 **c** 2

19 The finger provides less contact with the live wire and hence more resistance.

20 A fuse will melt when a high current flows in a circuit. Without the fuse the heat generated from a high current could be enough to start a fire and burn the house down. A safety switch switches off a circuit when the current in the active and neutral wires are not equal, thus preventing possible electrocution.

Unit 1 Area of Study 2 Review

1 B **2** A

3 **a** 1.3×10^{19} electrons **b** 1.60×10^{-17} J

c The electrical energy is converted into heat energy in the wire.

d $P = 2.0 \times 10^2$ W **e** 2.0×10^2 W **f** 2.0×10^2 W

g These answers are the same because power is the energy given to each unit of charge (volt).

4 Current only flows when there is a potential difference. If the bird touches only one wire the potential difference between the feet is negligible. If it touches two different wires they will be at different voltages, and so current will flow.

5 **a** Charge is escaping at the same rate.

b 1.25×10^{13} electrons **c** $P = 0.8$ W

d Energy = 6.40×10^{-14} J

6 **a** $R = 50\ \Omega$ **b** $U = 1.4 \times 10^{-18}$ J **c** $P = 1.6$ W

d The number of electrons per second = 1.1 × 1019

7 B **8** C

9 **a** $R_T = 60\ \Omega$ **b** $I = 2.0$ A

c $I_1 = 1.20$ A, $I_2 = 0.60$ A, $I_3 = 0.20$ A

d $P = 240$ W **e** 240 W

10 **a** $R = 6.7$ kΩ **b** $R_T = 2.5$ kΩ

11 **a** $R = 0.13\ \Omega$ **b** $V_{\text{drop}} = 1.3$ V **c** $P = 13$ W

12 **a** $P = 2.0$ W **b** $I = 0.67$ A

13 $I = 2.0$ A

14 **a** $R = 50\ \Omega$ **b** $R = 50\ \Omega$ **c** $R = 56\ \Omega$ **d** $R = 67\ \Omega$

15 $P = 405$ W

16 The resistance of the element increases with temperature.

17 **a** $I = 0.27$ A. We would not expect twice the current at 4 V. If the curve is extended to 4 V we could expect around 0.31 A, but it might well be that the globe would have burnt out before 4 V is reached, as the graph implies that it is probably a 3 V globe.

b This is not an ohmic conductor because the I–V graph is non-linear.

c $P = 0.9$ W

18 **a** The device is non-ohmic. The purpose of this device is to limit the current through a particular section of the circuit to a constant value regardless of the voltage across that part of the circuit.

b The resistance of the device increases with voltage.

c $V_2 = 100$ V $V_1 = 150$ V

d $P = 0.30$ W **e** $P = 0.20$ W **f** $P = 0.50$ W

19 Efficiency = 53% **20** $t = 2.8$ s **21** $R = 42$ A

22 $V = 0.25$ V **23** $V = 219$ V

24 power rating = 60 W; voltage rating = 60 V

25 Total current = 1.0 A, which is the same as for the parallel connection.

26 The power bill is calculated on the total power consumption, which is the same for both arrangements. Mary is correct.

27 $R_e = 4.3\ \Omega$ **28** **a** $V = 12$ V **b** $V = 8$ V **c** $R = 16\ \Omega$

29 A **30** D **31** C **32** A

33 **a** $R = 50\ \Omega$ **b** $R = 15\ \Omega$

34 **a** $I = 20$ mA **b** $I = 100$ mA **35** A **36** D **37** A

38 A fuse protects against overload current—too much current poses a fire hazard. An RCD detects an imbalance between current entering and leaving a device, which suggests that current is flowing to earth. Both will throw a circuit breaker.

39 Household circuits are connected in parallel, so that each device is supplied with 240 V and can be turned on and off individually. Within a single circuit all the current to the devices in parallel passes through the one circuit breaker.

40 To protect against drawing too much current and exceeding the rating, posing a fire risk.

41 $I = 3.33$ V **42** $R = 72\ \Omega$

43 Current drawn by circuit = 16.6 A. As the circuit has a 15 A circuit breaker, the circuit breaker will trip.

44 Total energy usage = 936 kWh **45** temperature = 20°C

46 Resistors R_1 and R_2 in the potentiometer form a voltage divider, which can set a variable voltage across the load. This in turn will control the current that flows through the load resistor.

47 Transducers convert one form of energy to another.
A microphone picks up the kinetic energy of air molecules and converts this to an electrical signal, and a speaker takes and electrical signal and converts it to motion of air molecules.

48 D

49 A short circuit is effectively bypasses the load in the circuit and connects the active and the neutral wires. This results in a greatly reduced resistance, and a high current flow. This condition will trigger the circuit breaker.

50 Plugs with three prongs have an earth. This is required when there is any possibility that the active lead could contact a metal and risk electrocuting a user. Some smaller devices are double insulated and so the active cannot deliver charge to any part of the device that a user can touch. In this case the earth is not needed and the plug can safely have only two prongs.

Chapter 5 The origins of everything

5.1 Measurements in the universe

WE 5.1.1 v = 17 500 km s^{-1}

1 0.38 **2** 4.0×10^{16} m **3** 4.2 ly

4 2.7×10^5 AU **5** 3.086×10^{25} m **6** 67 km s^{-1} Mpc^{-1}

7 500 Mpc **8** C

9 Hubble achieved this by measuring the redshift in the spectrum of the galaxies. **10** 24 500 km s^{-1}

5.2 The big bang

1 It is radiation from the early universe whose wavelength has expanded with the expansion of the universe.

2 The steady state theory suggests that the universe was infinite and that matter was being created all the time at just the right rate to keep the density constant as it expanded.

3 The CMB radiation was generated in the initial very hot state of the universe, which the steady state theory did not support. The steady state theory had no explanation for the CMB.

4 As the universe cooled, the energy of photons decreased which consequently decreased pair production and most particles annihilated with their antiparticles. However, there was a slight imbalance of matter and antimatter which left the universe mostly with matter particles.

5 The matter created during inflation would never have condensed to form atoms, and therefore, ultimately, galaxies or stars.

6 Annihilation, which involves the conversion of an electron and positron into two photons.

7 Although the initial radiation was largely the result of matter–antimatter annihilation at around 0.001 seconds, it continually interacted with the extremely hot nuclei for the next 300 000 years. It was only when the radiation had cooled below the temperature at which it was hot enough to ionise atoms that atoms could form and the radiation became stable and simply filled the universe. As the universe expanded so too did the wavelength of the radiation.

8 There was a tiny fraction of a percent more matter than antimatter created in the first microseconds of the universe.

9 **a** At that time hydrogen, helium and lithium were the only elements to exist. They had been formed in the first few minutes while the universe was hot enough for fusion to occur. Other elements didn't form as the universe was still hot enough that photons had enough energy to ionise any atoms that did form.

b Other elements formed from the supernovae of the early, large stars. These supernovae not only formed other stars, but eventually whole solar systems.

5.3 Particles of the Standard Model

1 between nucleons i.e. protons and neutrons

2 weak nuclear: W^+, W^- and Z bosons; strong nuclear: gluons; electromagnetic: photons.

3 The gauge bosons are the force-carrier particles. The leptons are fundamental particles that can be found individually and do not experience the strong force. The hadrons all experience the strong nuclear force and are made of quarks.

4 Quarks must exist in groups of two or three; leptons can exist individually.

5 The ball represents the force-carrier particle being exchanged, namely bosons.

6 The two groups of hadrons are called mesons and baryons. The mesons contain two quarks. One of the quarks in this group is normal matter while the other is antimatter. The baryons have three quarks. This group contains the familiar particles called protons and neutrons.

7 All quarks experience the strong force; leptons do not. Quarks also have non-integer (fractional) charges; leptons have charges of −1 or 0.

8

Gauge boson	Lepton	Hadron
gluon, photon	electron, neutrino, muon	neutron, proton

Chapter 5 Review

1 The light from a source that is moving away from the observer will appear to have a longer wavelength and a lower frequency.

2 Galaxy 3 has the greatest redshift, then galaxy 2 and galaxy 1 has the least redshift. Therefore in order of increasing distance from Earth they are: galaxy 1, galaxy 2, galaxy 3.

3 There is a relationship between the period of the variation in their brightness and their intrinsic brightness. It is possible to work out the distance to these stars by comparing their intrinsic brightness to their apparent brightness.

4 For each Mpc a galaxy is away from us, it is receding from us at 67.80 km s^{-1}.

5 A conventional explosion has a definite centre or source and the shock wave expands into the space around it. In the expansion of the universe, there is no centre to the expansion and as the universe is expanding space is also expanding; the universe does not expand 'into' space.

6 For a steeper line of best fit, the Hubble constant would be greater and this would indicate the universe was younger than Hubble would have calculated from the graph.

7 14 billion years **8** 3.3×10^8 years

9 fusion; extreme; did not; billions; gravitational; heavier

10 As a result of the expanding universe, the CMB radiation will be detected at a longer wavelength, which corresponds to a lower frequency and less energy.

11 A photon creates a particle and its antiparticle and energy is converted into mass.

12 A particle collides with its antiparticle and mass is converted into energy.

13 protons: up, up, down; neutrons: up, down, down

14 electromagnetism and the strong and weak nuclear forces

15 A proton is made up of two up quarks ($2 \times +\frac{2}{3}$) and one down quark ($-\frac{1}{3}$) so $\frac{4}{3} - \frac{1}{3} = +1$.

A neutron is made up of two down quarks (2 × –1/3) and one up quark (+2/3) so –2/3 + 2/3 = 0.
16 weak nuclear, electromagnetic, gravity, (strong nuclear)
17 The Standard Model is based on the assumption that forces arise through the exchange of particles called gauge bosons (or just bosons). Each of the three forces is mediated by a different particle: strong—gluon, electromagnetic—photon, weak—W^+, W^- and Z.
18 fundamental particle; lepton

Chapter 6 Particles in the nucleus

6.1 Atoms, isotopes and radioisotopes

WE 6.1.1 90 protons, 230 nucleons and 140 neutrons
1 nucleons **2** 79 protons and 118 neutrons **3** 235
4 **a** 17 protons, 18 neutrons and 35 nucleons
  b 94 protons, 145 neutrons and 239 nucleons
5 B and D
6 It is the same as the number of protons, which is given by the atomic number.
7 Isotopes are atoms with the same number of protons but different numbers of neutrons.
8 **a** Their atomic numbers are the same as they are both krypton. Their mass numbers (84 and 89) are different as they are isotopes.
  b There would be no difference in their chemical interactions with other atoms.
9 A radioisotope is an unstable isotope. At some time, it will spontaneously eject radiation in the form of alpha particles, beta particles or gamma rays from the nucleus.
10 Yes, a natural isotope can be radioactive. For example, uranium is naturally occurring and every isotope of uranium is radioactive.

6.2 Radioactivity

WE 6.2.1 mass number = 214, atomic number = 82, lead
1 X is an alpha particle.
2 Y is a beta minus particle
3 beta plus
4 a positively charged electron
5 Alpha is a helium nucleus. Beta is a positively or negatively charged electron. Gamma is electromagnetic radiation.
6 40, 42, 43, 44, 46, 48
7 Alpha, beta and gamma radiation all originate from the nucleus of an atom.
8 **a** X: atomic number = 90, mass number = 231, X is thorium
  b Y: atomic number = 89, mass number = 228, Y is actinium
9 **a** 7 protons and 7 neutrons
  b A neutron has changed into a proton, an electron and an antineutrino.
10 **a** beta minus particle **b** alpha particle

6.3 Properties of alpha, beta and gamma radiation

1 **a** gamma **b** beta minus **c** alpha **d** beta **e** gamma
2 gamma **3** beta **4** gamma
5 beta and gamma **6** **a** nucleus **b** nucleus **c** nucleus
7 gamma, beta, alpha
8 Alpha particles travel through air at a relatively low speed and have a double positive charge. Their charge and their relatively slow speed make them very easy to stop. This means that they have a very poor penetrating ability.
9 The wire should be a beta emitter, since the irradiation needs to be confined to a relatively small area. Alpha radiation does not have sufficient penetrating power, while gamma radiation would irradiate adjacent healthy cells.
10 Alpha particles will all be stopped by the metal sheet. Gamma rays will all penetrate the metal sheet. Differences in the thickness of the metal sheet will not affect the count rates of these two. Some beta particles will pass through thin metal so for a set metal thickness there is a set rate of beta that should make it to the other side.

6.4 Half-life and decay series

WE 6.4.1 3.9×10^7
1 the count rate or the number of decays each second
2 $N = 4.0 \times 10^{10}$ **3** $N = 3.0 \times 10^{11}$
4 **a** 10 half-lives **b** 240 000 years **5** 50%
6 192 μg must be produced
7 the half-life of the radioisotope is 15 minutes
8 **a** 10 minutes **b** 50 Bq
9 beta decay; 20 years
10 Seven alpha and four beta-minus decays have occurred.

Chapter 6 Review

1 20 protons and 25 neutrons
2 27 protons, 33 neutrons and 60 nucleons
3 X is a gamma ray.
4 beta minus
5 **a** beta minus **b** proton **c** alpha **d** neutron **e** gamma
  f beta positive (positron)
6 atomic number = 3, mass number = 7, so X is lithium
7 **a** X is a proton **b** Y is a neutron
8 **a** $x = 208$ $y = 82$ **b** $x = 176$ $y = 78$
9 $a = 18$ $b = 9$ $c = 18$ $d = 8$
  X is fluorine, F. Y is oxygen, O.
10 X is carbon-12
11 Electromagnetic forces are balanced by the strong nuclear force acting between all nucleons in close proximity.
12 **a** gamma **b** gamma **13** gamma radiation
14 The bombarding electrons will be strongly repelled by the electron clouds of the atoms as they are all negatively charged. The small mass of the bombarding electrons also makes them relatively easy to repel compared to, say, a proton.
15 $N = 3.0 \times 10^{14}$ **16** $N = 7.0 \times 10^{14}$
17 Uranium-235 has a greater activity.
18 **a** 2 MBq **b** 6 hours **c** 5.0×10^5 Bq
19 $N = 1.5 \times 10^{10}$
20 The long half-life means that the source will not need to be replaced for many years. The gamma rays have a strong penetrating power so they are able to penetrate the skull and reach the tumour site.

Chapter 7 Energy from the atom

7.1 Nuclear fission and energy

WE 7.1.1 **a** 3, **b** 4.12×10^{-11} J = 258 MeV, **c** 0.16%
1 The strong nuclear force is a force of attraction that acts between every nucleon but only over relatively short distances. This force acts like a nuclear cement.
2 The decay products of the nuclear fission process comprise many different, often highly radioactive isotopes. This is what makes up the waste.
3 Since the neutron is neutral it will only experience attractive forces from other nucleons.
4 8.0×10^{-13} J **5** 3.8×10^4 eV
6 Fissile—uranium-235 and plutonium-239; non-fissile—uranium-238 and cobalt-60.
7 $x = 3$ **8** **a** $E = 1.91 \times 10^{-11}$ J **b** $E = 1.19 \times 10^8$ eV

7.2 Nuclear fusion

1 Fusion is the joining together of two small nuclei to form a larger nucleus. Fission is the splitting apart of one large nucleus into smaller fragments.
2 The mass of the products is less than the mass of the reactants. The mass difference is related to the energy released via $E = mc^2$.
3 The amount of energy released per nucleon during a single nuclear fission reaction is less than the amount for a single fusion reaction.
4 less than 1%
5 **a** $a = 4$ $b = 2$ X is helium, He **b** $m = 5.9 \times 10^{-29}$ kg
6 Electrostatic forces of repulsion act on the protons. If the protons are moving slowly they will not have enough energy to overcome the repulsive forces and they will not fuse together.

7 Initially, electrostatic forces of repulsion act on the protons, but they are travelling fast enough to overcome these forces. The protons will get close enough for the strong nuclear force to take effect and they will fuse together. These protons have overcome the energy barrier.

8 **a** Atomic number = 2, mass number = 3, particle X is $^3_2$He
b $E = 3.7 \times 10^{-12}$ J **c** $m = 4.1 \times 10^{-29}$ kg

9 When two hydrogen-2 nuclei are fused together to form a helium-4 nucleus, the binding energy per nucleon increases and the nucleus becomes more stable.

10 The number of nucleons is conserved as there are five nucleons on each side of the reaction.

7.3 Electromagnetic waves and synchrotron radiation

1 Electromagnetic radiation covers a wide range of frequencies (or wavelengths) of rapidly changing electric and magnetic fields that can be created by accelerating charges and are travelling away from the source at the speed of light.

2 The wavelength decreases **3** the frequency of the radiation

4 approximately 400–800 nm **5** from microwaves to X-rays

6 by electrons that are accelerated around a curved path

7 They all travel at the speed of light, 3×10^8 m s–1 or 300 000 km s^{-1}. **8** 1.6×10^5 eV

9 **a** linac—to accelerate electrons using powerful bursts of RF radiation
b booster ring—to accelerate electrons further as they pass through RF cavities until reaching an energy of 3 GeV
c storage ring—to accelerate electrons in a curved path steered by the strong magnetic fields of the dipole magnets and insertion devices
d beamline—to carry synchrotron radiation from the storage ring to an experimental station

10 Synchrotron radiation travels faster (at the speed of light) than the electron beam.

7.4 The production of light

WE 7.4.1 0.42 eV

1 An emission spectrum for an element typically consists of a series of spaced coloured lines on a black background.

2 The different coloured lines in the emission spectrum of an atom correspond to the possible electron transitions between energy levels within the atom.

3 The energy levels within an atom are commonly represented as horizontal lines on a graph.

4 the ground state

5 No. The photon energy will be exactly equal to the energy difference between the electron's initial and final levels.

6 The light globe produces a continuous spectrum showing all the colours of the rainbow. The vapour lamp produces a discrete spectrum showing just coloured lines.

7 Electron transitions between energy levels in sodium atoms produce photons of the correct wavelength (or frequency) that correspond to orange light.

8 There are three possible ways the electron can return from the $n = 3$ level to the ground state: $n = 3$ direct to $n = 1$, $n = 3$ to $n = 2$ and $n = 2$ to $n = 1$. A different amount of energy is released for each transition.

9 **a** $n = 1$ **b** $n = \infty$ **c** $n = 2$, $n = 3$ and $n = 4$ **10** 10.2 eV

Chapter 7 Review

1 A nuclide that is able to split in two when hit by a neutron is fissile.

2 No, only a few nuclides (e.g. uranium-235 and plutonium-239) are fissile.

3 The strong nuclear force causes the proton to be attracted to all other nucleons. It will also experience a smaller electrostatic force of repulsion between itself and other protons.

4 Neutrons are uncharged and are not repelled by the nucleus as alpha particles are.

5 **a** 3.1×10^{-11} J **b** 196 MeV

6 $x = 239$ $y = 40$ **7** $x = 5$

8 The nuclei are all positively charged and so repel each other. They need a massively large amount of energy to overcome these forces and get close enough for the strong nuclear force to take effect. 100 million degrees provides the required energy for this to occur.

9 4.49×10^{-11} J

10 **a** The combined mass of the hydrogen and helium-3 nuclei is greater than the combined mass of the helium-4 nucleus, positron and neutrino.
b The energy has come from the lost mass (or mass defect) via $E = mc^2$.
c 3.4×10^{-12} J **d** $m = 3.8 \times 10^{-29}$ kg

11 Fission reactors create a great deal more waste. Fusion creates more energy per nucleon than fission.

12 The binding energy per nucleon increases and the nucleus becomes more stable.

13 The higher the binding energy, the more stable the nucleus. This is because higher binding energy means that it takes more energy to completely separate particles in the nucleus. Iron therefore has the most stable nuclei of all the elements.

14 The temperature of the light globe filament *increases* when it is switched on. This causes *charged* particles in the filament to vibrate (or accelerate) and some *visible light* and a lot of *infrared* radiation are produced.

15 8.0×10^{-15} J **16** gamma rays **17** beta rays

18 a tangent to the curved path of the electrons

19 2.11 eV **20** 3.61 eV

21 There are *three* different energies: 6.7 eV, 4.9 eV and 1.8 eV.

Unit 1 Area of Study 3 Review

1 The big bang is an expansion of space–time. Before the big bang there was no space, time or matter, so it is not a case of matter exploding out into space in a time continuum, but space and time itself being created at the big bang event as energy converted to matter, after which space expanded. The energy present allowed the creation of matter–antimatter pairs and the rapid inflation of the universe prevented annihilation taking place immediately, taking the created matter with it. While it is true that the early universe was extremely dense, the big bang theory would suggest that mass/energy, space and time all emerged at once from nothing.

2 Yes, he is correct in the sense that the light that reaches our telescope left the object viewed at some previous point in time. The more distant the object, the longer the light takes to reach us, and in a sense the further back in time we are looking.

3 Stellar spectra include the absorption lines for elements for which we know the wavelengths. When these spectra are emitted from objects receding from us, the lines are shifted to longer wavelengths—towards the red end of the spectrum. This is the case for the majority of stellar objects which are redshifted, the more distant objects being shifted the most and hence receding the fastest. This is consistent with a model in which space–time and the whole universe is expanding.

4 6300 km s^{-1}, or 2% of c

5 **a** Recession velocity of Proxima Centauri is 91 mm s^{-1}
Recession velocity of edge of the universe is 9.8×10^8 m s^{-1}
b The edge of the visible universe is receding from us at a speed in excess of the speed of light. That is not a violation of the principles of special relativity, as no object is moving through *space* at a speed in excess of the speed of light, it is purely a relative velocity.
If there is concern that technically this is not the edge of the visible universe, since light would never reach us, we are only concerned here with a factor of 3 or so and in astronomical terms, this is good enough!
c $H_0 = 2 \times 10^{-18}$ s^{-1}
The relative proximity of a star like Proxima Centauri would mean redshift would not be measurable.

d The extremely small value for H_0 reflects the fact that the recession velocity is only significant for huge distances.

6 Cepheid variables have a period that varies with their absolute brightness. So by observing their period, it is possible to infer their absolute brightness. Their actual brightness can be measured, and so it is possible to work out their actual distance by using the drop in intensity of the radiation.

7 **a** The radiation which when created would have had a very short wavelength would be expected to 'stretch out' with space itself, and so would have a much longer wavelength as space expanded. Calculations show that this would be in the microwave range today.

b The variations indicate a slightly uneven distribution of light and therefore matter. This allowed gravitational attraction to collect clumps of matter together, ultimately forming stars and galaxies. If there had been completely uniform radiation, there would have been no universe as we know it.

8 Pair production is the creation of a matter and antimatter pair of particles, such as a positron and an electron from a photon. This is a mechanism for the creation of particles from photons.

9 Normally pairs annihilate rapidly with the release of photons, but the rapid inflation moved the pairs apart so that the particles were able to persist.

10 As the universe cooled, the average photon energy dropped to a level at which a photon no longer had the energy required to create a matter–antimatter pair

11 Any atoms formed would immediately be ionised as the photons, although not having enough energy for pair production, certainly possessed the ionisation energy for a hydrogen atom.

12 Fusion requires very high densities, temperatures and pressures for charged particles overcome their mutual repulsion and come close enough for the strong nuclear force to exceed the electrostatic repulsion. This happened in the first few seconds after the big bang, and then particle distances increased and energies dropped below the values required for fusion to be possible. Fusion re-ignited in stars much later when gravitational forces once again brought particles together at high densities.

13 Photon energies had to be below the ionisation energy of the atoms.

14 Gravity caused particles to aggregate. As the dust clouds collapsed under their mutual attraction, vast amounts of energy were released and this created the temperatures and pressures for fusion to re-ignite in the first stars.

15 **a** The energy of photons results from the mass of the leptons being converted to energy.

b 1.6×10^{-13}

16

Particle	Property
gluon	mediator of the strong nuclear force interacts with quarks
photon	mediator of the electromagnetic force interacts with charged particles
W^+, W^- and Z	mediator of the weak nuclear force causes nuclear decay
(graviton)	mediator of the gravitational force

17

Category	Particles	Description
BOSONS	photons, gluons, gravitons, W+, W-, and Z	mediators of the fundamental forces
FERMIONS make up all matter	leptons: positrons, electrons neutrinos, muons	• do not experience the strong force • exchange W and Z bosons—weak nuclear force • charged leptons exchange photons—electromagnetic force
	hadrons	do experience the strong force, exchanging gluons made up of quarks
	1 baryons: protons, neutrons, antiprotons	made of 3 quarks
	2 mesons: pions	made of 2 quarks

18 **a** The typical nucleon separation lies at the point where the attraction of the strong nuclear force is the strongest. This is the short-range force that binds the nucleons to their nearest neighbours.

b The strong force acts on protons and neutrons within range. The elctrostatic force acts between protons.

c For larger atoms, the repulsive forces between protons accumulate. The electrostatic force causes protons to repel one another and this accumulates as proton number increases. This destabilises the nucleus. Neutrons supply additional binding as they introduce the strong force without the cost of repulsion, thereby increasing stability.

19 Cs-137: 55 protons, 82 neutrons, 137 nucleons
I-131: 53 protons, 78 neutrons, 131 nucleons

20 **a** nucleus **b** nucleus **c** nucleus

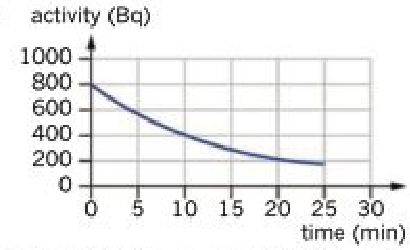
21

Time (× 10^9 years)	No. of K nuclei	No. of Ar nuclei	Ratio K:Ar
0	1000	0	–
1.3	500	500	1:1
2.6	250	750	1:3
3.9	125	875	1:7

22 ${}^{7}_{3}\mathrm{Li} + {}^{1}_{1}\mathrm{H} \rightarrow 2\,{}^{4}_{2}\mathrm{He}$

23 **a** 8.8 MeV **b** 0.42 MeV **c** 0.5 MeV

24 **a**



b about 320 Bq **c** about 10 min **d** about 100 Bq

25 B

26 **a** 1 p, 2 n **b** helium **c** B

d A neutron has transformed into a proton and an electron.

27 ${}^{185}_{79}\mathrm{Au} \rightarrow {}^{181}_{77}\mathrm{Ir} + {}^{4}_{2}\mathrm{He}$

28 **a** 1 minute **b** 4.7 g **c** ${}^{11}_{16}\mathrm{Na} \rightarrow {}^{0}_{-1}\beta + {}^{24}_{12}\mathrm{Mg}$ + energy

29 **a** B

b The activity of Bi-211 is 4 times greater at the start, but after 8 minutes (4 half-lives) its activity is reduced by a factor of 16. After 8 minutes the activity of Bi-215 has fallen by half and so it has double the activity of Bi-211 at this time.

30 **a** ${}^{238}_{92}\mathrm{U} \rightarrow {}^{234}_{90}\mathrm{Th} + {}^{4}_{2}\mathrm{He}$ **b** ${}^{218}_{84}\mathrm{Po} \rightarrow {}^{218}_{85}\mathrm{At} + {}^{0}_{-1}\beta$

31 Po and At have different numbers of protons and this is what makes them distinct elements.

32 ${}^{210}\mathrm{Bi}$ can undergo beta decay to form ${}^{210}\mathrm{Po}$ and then this undergoes alpha decay to form ${}^{206}\mathrm{Pb}$. Alternatively, it an undergo an alpha decay first to form ${}^{206}\mathrm{Tl}$ and then the subsequent beta decay results ion ${}^{206}\mathrm{Pb}$.

33 They all have 84 protons, but differ in their number of neutrons. 214, 210 and 206 neutrons respectively

34 The total energy that is released when the nucleons bind to create the nucleus, divided by the number of nucleons. This is then also effectively the energy required to break up the nucleus into its constituent nucleons.

35 Fe requires the most energy per nucleon to break up the nucleus, therefore it is the most stable.

36 The energy per nucleon for uranium is about 7.5 MeV and the binding energy per nucleon for fragments of mass number 118 is 8.5 MeV. That means that when the smaller fragments are formed, they are more tightly bound and the difference in energy is released in the fission reaction. This is about 2 MeV for each nucleon.

37 470 MeV

38 Hydrogen is an unbound proton, so it has no binding energy.

39 The nucleons release the binding energy as a result of becoming bound in the He nucleus.

40 2.7×10^3 tonnes of coal

41 a There is approximately the same number of protons and neutrons in small stable nuclei.

b There are more neutrons than protons in large unstable nuclei.

c The strong nuclear force which binds the nucleons together is greater than the repulsive electrostatic force acting on the proton as a result of the other 5 protons, hence the nucleus is stable.

d The repulsive electrostatic force on the proton is greater because of the increased charge concentrated in the nucleus, but the strong nuclear force which only acts on neighbouring nucleons is of the same order. This nucleus is therefore less stable.

42 a The reactants have greater combined mass than the fusion products.

b kinetic energy gained by the fusion products and electromagnetic radiation

c As the nuclides approach, an electrostatic force of repulsion is acting on them, but their great speed enables them to overcome this force and fuse together. At this point, electrostatic forces of repulsion and larger nuclear forces of attraction are acting on the protons. The neutrons experience only strong nuclear forces of attraction.

43 a 4.49×10^{-11} J **b** 280 MeV

44 a Electrostatic forces of repulsion act on the protons. They do not have enough energy to overcome this force to get close enough for the strong nuclear force to come into effect and hence will not fuse. These protons have not jumped the energy barrier.

b Electrostatic forces of repulsion act on the two protons initially, but the protons have enough energy to push past these forces and get close enough together for the strong nuclear forces to take effect. This force enables the nucleons to fuse. These protons have overcome the energy barrier.

45 Fusion joins lighter elements to form a heavier element and fission splits a heavy element to form lighter fragments.
Fusion usually requires the reacting nucleons to have high energy to overcome the electrostatic repulsion. Fission can be achieved with the capture of relatively slow-moving neutrons.
Fusion does not create harmful and radioactive waste products but fission wastes are often highly radioactive.
At this point fission has been achieved commercially, but fusion reactors are still under development.
Energy released per nucleon is higher for fusion than for fission.

46 a Accelerating charges are caused by thermal vibration of electrons in the hot filament.

b Electrons are decelerating as they are fired into metal targets.

c Electrons are accelerated between two electrodes and excite the electrons in a sodium vapour. The electrons jump up to higher energy levels and then emit the characteristic yellow light corresponding to the dominant electron transitions for sodium.

d Thermal vibrations cause acceleration of the charges. Because the steel is at a lower temperature than the tungsten, it emits at a longer wavelength.

e Electrons accelerated around the circular beamline radiate.

f Electrons are pumped up into higher metastable energy levels and then are stimulated to drop down to a lower level.

Chapter 8 Scalars and vectors

8.1 Scalars and vectors

WE 8.1.1 **a** 50 N west, **b** –50 N

WE 8.1.2 50° clockwise from the right direction

1 Scalar measures require a magnitude (size) and units.

2 Vectors require a magnitude, units and a direction.

3

Scalar	Vector
time	force
distance	acceleration
volume	position
speed	displacement
temperature	momentum
	velocity

4 a 5.4 N **b** 2.7 N **c** 8.1 N

5 a 10.8 N **b** –5.4 N **c** 16.2 N

6 a down **b** south **c** forwards
d up **e** east **f** positive

7 Terms like north and left cannot be used in a calculation. + and – can be used to do calculations with vectors.

8 –35 N

9 a i 225° T and **ii** S 45° W **b i** 120° T and **ii** S 60° E

10 40° clockwise from the left direction

8.2 Adding vectors in one and two dimensions

WE 8.2.1 19 N down **WE 8.2.2** 5.8 N, N 59° E

1 4 m west **2** 2 m down **3** forwards **4** D

5 R = 44.7 m, S 63.4° W **6** 6325 N, N 71.6° E **7** C

8.3 Subtracting vectors in one and two dimensions

WE 8.3.1 1988 m s^{-1} up **WE 8.3.2** 9.2 m s^{-1} N 41° E

1 8 m s^{-1} east **2** 2 m s^{-1} left

3 7 m s^{-1} downwards **4** 67.5 m s^{-1} south

5 14.0 m s^{-1} backwards **6** 533 m s^{-1} N 49.6° W

7 59.4 m s^{-1} N 45.0° W **8** 8.79 m s^{-1} N 36.7° W

8.4 Vector components

WE 8.4.1 3168 N left, 1580 N downwards

1 a F_D = 265 N downwards **b** F_R = 378 N right

2 19.8 N south, 16.6 N east **3** 4.55 ms^{-1} north, 17.7 ms^{-1} west

4 The student is 18.9 m south and 43 m east of his starting point.

5 109 000 N north, 208 000 N east

6 a 50 N south, 87 N east **b** 60 N north,
c 280 N south, 103 N east **d** 1.5×10^5 N up, 2.6×10^5 N horizontal

7 horizontal component F_h = 150 N
vertical component F_v = 260 N

8.5 Mass and weight

1 50 kg **2** 60 kg is the student's mass, not her weight.

3 735 N **4** 3.5 kg **5** 5.6 N

6 The mass of the hammer remains constant at 1.5 kg.
The weight of the hammer on Mars is 5.4 N.

7 The weight of any object will be less on the Moon compared with its weight on the Earth as gravity is weaker on the Moon, due to its smaller mass.

Chapter 8 Review

1 B and D **2** A and D

3 The vector must be drawn as an arrow with its tail at the point of contact between the hand and the ball. The arrow poiOnts in the direction of the 'push' of the hand.

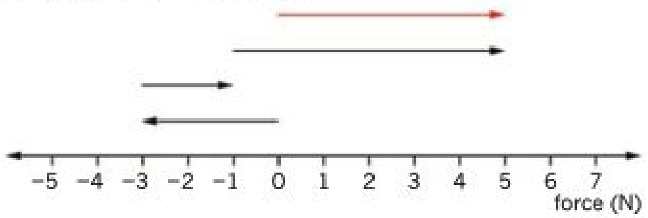
4 A has twice the magnitude of B.

5 Signs are useful in mathematical calculations, as the words north and south cannot be used in an equation.

6 34.0 m s^{-1} north and 12.5 m s^{-1} east. **7** + 80 N or just 80 N

8 70° anticlockwise from the left

9



The resultant vector is 5 N right.

10 21.0 m backwards 11 65.7 m S 56.8° W
12 813 N, N 53.7° E 13 6 m s^{-1} left
14 22.8 m s^{-1} N 55.2° W 15 67.7 m s^{-1} N 35.0° W
16 39.4 N east, 22.8 N south 17 98 N
18 2.1 kg 19 a 85 kg b 85 kg c 306 N down
20 Earth, Mars, the Moon

Chapter 9 Linear motion

9.1 Displacement, speed and velocity

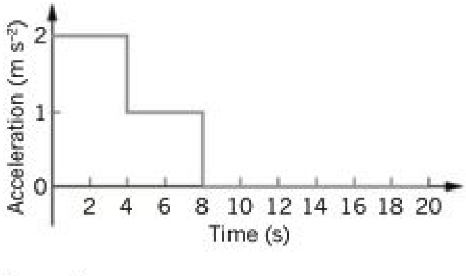
WE 9.1.1 **a** 0.92 m s^{-1} east, **b** 3.3 km h^{-1}, **c** 4.0 m s^{-1}, **d** 14.4 km h^{-1}
1 B and C
2 a displacement = +40 cm, distance travelled = 40 cm
b displacement = –10 cm, distance travelled = 10 cm
c displacement = 20 cm, distance travelled = 20 cm
d displacement = 20 cm, distance covered = 80 cm
3 a 80 km b +20 km or 20 km north
4 a –10 m or 10 m downwards b +60 m or 60 m upwards
c 70 m d 50 m or 50 m upwards
5 a 33 m s^{-1} b 25 m 6 a 17 km h^{-1} b 4.7 m s^{-1}
7 a 0.9 m s^{-1} b 0.1 m s^{-1} east
8 a 10 km h^{-1} b 2.8 m s^{-1} south
9 a 21 km b 15 km north c 14 km h^{-1} d 10 km h^{-1} north

9.2 Acceleration

WE 9.2.1 **a** –2.0 m s^{-1}, **b** 16 m s^{-1} up **WE 9.2.2** 460 m s^{-2} up
1 –7 km h^{-1} 2 +5 m s^{-1} or 5 m s^{-1} up
3 +8 m s^{-1} or 8 m s^{-1} up 4 –5.0 m s^{-2} or 5.0 m s^{-2} south
5 43 km h^{-1} s^{-1}
6 a –10 m s^{-1} b –40 m s^{-1} or 40 m s^{-1} west c 800 m s^{-2}
7 a 8.0 m s^{-1} b –8.0 m s^{-1} or 8.0 m s^{-1} south c 6.7 m s^{-2}

9.3 Graphing position, velocity and acceleration over time

WE 9.3.1 **a** –15 m s^{-1} or 15 m s^{-1} backwards, **b** The cyclist is not moving and the velocity is 0 m s^{-1}.
WE 9.3.2 **a** 4 m west **b** 2 m s^{-1} west **WE 9.3.3** 2 m s^{-2} west
1 D
2 The car initially moves in a positive direction and travels 8 m in 2 s. It then stops for 2 s. The car then reverses direction for 5 s, passing back through its starting point after 8 s. It travels a further 2 m in a negative direction before stopping after 9 s.
3 a +8 m b +8 m c +4 m d –2 m 4 at t = 8 s
5 a +4 m s^{-1} b 0 c –2 m s^{-1} d –2 m s^{-1} e –2 m s^{-1}
6 a 18 m b –2 m
7 a 5 m s^{-1} b 20 m s^{-1} north c 10 m s^{-1} north
8 a 0 m s^{-2}
b –1 or just 1 m s^{-2}
Magnitude only so direction is not required.
c 10.5 m d 1.5 m s^{-1}
9 a 20 m s^{-1} north b –40 or 40 m s^{-1} south
10 a 80 s b approx 1.2 or 1.3 m s^{-2}
c approx 0.38 or 0.41 m s^{-2} d 4900 m or 4.9 km
11 a 2 m s^{-1} b t = 10 s c 80 m d 7 m s^{-1}
12 a



b +12 m s^{-1}

9.4 Equations for uniform acceleration

WE 9.4.1 **a** 3.8 m s^{-2} west, **b** 3.9 s, **c** 7.5 m s^{-1} east
1 E 2 a 3.1 m s^{-2} b 50 m s^{-1} c 180 km h^{-1}
3 a 2.0 m s^{-2} b 8 m s^{-1} c 64 m
4 a 40 m s^{-2} b 1.12 or 1.1 km c 576 or 580 km h^{-1}
d 80 m s^{-1} e 124.4 or 120 m s^{-1}
5 a 5.0 m s^{-2} b 2.725 or 2.7 m c 5.45 or 5.5 m s^{-1}
6 D 7 a 98 m s^{-2} b 0.29 s c 19.8 or 20 m s^{-1}
8 a 20.83 or 21 m s^{-1} b 5.25 or 5.3 m
c 36.75 or 37 m d 42.3 or 42 m
9 a 4.0 m s^{-1} b 5.7 m s^{-1} c 2.0 s d 2.85 s
The time taken to travel the final 4.0 m is
2.85 s – 2.0 s = 0.85 s.
10 a t = 8.0 s b t = 16 s c 192 m

9.5 Vertical motion

WE 9.5.1 **a** 2.5 s **b** 3.5 s **c** –34 m s^{-1} or 34 m s^{-1} downwards
WE 9.5.2 **a** 11.5 m **b** 3.0 s
1 B 2 A and D
3 a 29 m s^{-1} b 24 m s^{-1} c 12 m s^{-1} (down)
4 a the same as b the same as
5 a 14.7 or 15 m s^{-1} b 11.25 or 11 m
6 a –3.92 or –3.9 m s^{-1} b –0.78 or 0.78 m
c –0.20 or 0.20 m d 0.58 m
7 a 2.0 s b 19.6 or 20 m s^{-1}
c 19.6 or 20 m d 20 m s^{-1} downwards
8 a 3.5 s b 2.9 s 9 a 1.7 s b 3.2 s

Chapter 9 Review

1 26 m s^{-1} 2 54 km h^{-1} 3 15 km h^{-1}
4 a 10 km h^{-1} north b 2.8 m s^{-1} north 5 –2.0 m s^{-1}
6 B 7 –6 m s^{-2}
8 a from 10 to 25 s b from 30 to 45 s
c from 0 to 10 s, from 25 to 30 s and from 45 to 60 s
d at 42.5 s or 43 s
9 a B b A c C
10 a 114 m north b 10.4 m s^{-1} b 0 m s^{-2}
d –7 or 7 m s^{-2} south e A
11 15.75 or 16 m s^{-1}
12 a 4.0 m s^{-2} b 4.0 m s^{-1} c 6.0 m
13 a –5.0 m s^{2} b 2.0 s
14 a +4 m b A and C c B, +0.8 m s^{-1}
d D, 2.4 m s^{-1} e 0.8 m s^{-1}
15 The marble slows down by 9.8 m s^{-1} each second so it will take 4 s to stop momentarily at the top of its motion. It has a positive velocity that changes to zero on the way up. Its acceleration is constant at –9.8 m s^{-2} due to gravity.
16 D 17 B
18 a 45 m b 6 s c –20 m s^{-1} or 20 m s^{-1} down
d 10 m s^{-2} down
19 a 10 s b 40.4 or 40 m s^{-1} c 6.7 s
20 15 m s^{-1} up 21 11 m

Chapter 10 Momentum and force

10.1 Newton's first law

1 The box has changed its velocity so an unbalanced force must have acted on the box to slow it down.
2 The car's direction has changed, which means the velocity has changed. An unbalanced force has acted on the car to change its direction.
3 B
4 No horizontal force acts on the person. The bus slows, but the standing passenger will continue to move with constant velocity unless acted on by an unbalanced force.
5 20 N
6 a 25 N b 25 N c 29 N at an angle of 30° to the horizontal.
7 They would have no retarding forces acting and so would tend to maintain their original velocity and move towards the front of the plane.
8 a Gravitational force of attraction between the two masses.
b Electrical force of attraction between the negative electron and the positive nucleus.
c Friction between the tyres and the road.
d Tension in the wire.

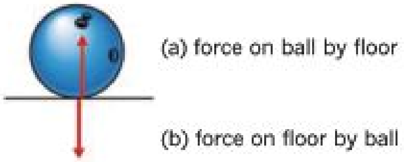
9 **a** If the cloth is pulled quickly, the force on the glass acts for a short time only. This force does not overcome the tendency of the glass to stay where they is, i.e. its inertia.
  b Using a full glass makes the trick easier because the force will have less effect on the glass due to its greater mass. The inertia of the full glass is greater than that of an empty glass.
10 The fully laden semitrailer will find it most difficult to stop. Its large mass means that more force is required to bring it to a stop.
11 lift = 50 kN up, drag = 12 kN west

10.2 Newton's second law

WE 10.2.1 307 N south **WE 10.2.2** 4.23 m s^{-1} left
WE 10.2.3 2.50 m s^{-2} forwards
WE 10.2.4 **a** 7.0 m s^{-2} forwards **b** 5.0 m s^{-2} forwards
1 6.61 m s^{-2} north **2** 38.3 kg
3 3.56 m s^{-2} north **4** 9.80 m s^{-2} down
5 9.8 m s^{-2} down **6** 0.547 m s^{-1} east
7 1560 m s^{-2} south **8** 78 500 kg
9 0.288 m s^{-1} north
10 **a** 50 kg **b** 490 N **c** 20 N south **d** 0.31 m s^{-2}
11 **a** 1.6 m s^{-2} **b** 0.8 m s^{-1} **c** 0.2 m s^{-2}
12 4 boxes **13** 10.2 m s^{-2}

10.3 Newton's third law

WE 10.3.1



1 There is a force on the hammer by the nail, and a force on the nail by the hammer. These two forces are equal in magnitude and opposite in direction.
2 **a** $F_{\text{on the astronaut by the Earth}}$ **b** $F_{\text{on the Earth by the astronaut}}$
3 The force on the hand by the water.
4 The force on the balloon by the escaping air. **3** 100 N east
6 **a** 140 N in the opposite direction to the leaping fisherman
  b 3.5 m s^{-2} in the opposite direction to the fisherman
  c speed of the fisherman = 1.0 m s^{-1}
  speed of the boat = 1.8 m s^{-1}
7 in the opposite direction from the ship
8 Tania is correct. For an action–reaction pair, the action force is a force on object A by object B, and the reaction force is a force on object B by object A. That is, the two forces act on *different* objects. In this case, both the weight force and the normal force are acting on the same object: the lunch box.

10.4 Momentum and conservation of momentum

WE 10.4.1 20 500 kg m s^{-1} north **WE 10.4.2** 0.80 m s^{-1} north
WE 10.4.3 1.00 m s^{-1} south **WE 10.4.4** 1630 m s^{-1} south
1 8.75 kg m s^{-1} south **2** 9610 kg m s^{-1} west
3 3.97 kg m s^{-1} south
4 The second ball has the greater momentum.
5 0.438 m s^{-1} backwards or –0.438 m s^{-1}
6 70.4 m s^{-1} **7** 4200 kg
8 3.0 m s^{-1} in the direction opposite to that of the exhaust gases
9 **a** v_2 = 40 m s^{-1} **b** 4.5 × 10^3 N **c** 10 m s^{-2}

10.5 Momentum transfer

WE 10.5.1 803 kg m s^{-1} south **WE 10.5.2** 0.0208 kg m s^{-1} N 38.7° E
1 83.1 kg m s^{-1} south **2** 235 000 kg m s^{-1} east
3 40.0 kg m s^{-1} east **4** 2.45 kg m s^{-1} down
5 2.4 m s^{-1} north **6** 2860 kg m s^{-1} N 45° E
7 377 kg m s^{-1} S 42° W

10.6 Momentum and net force

WE 10.6.1 **a** 0.192 kg m s^{-1} up **b** 0.192 kg m s^{-1} up **c** 54.1 N up
WE 10.6.2 **a** 0.192 kg m s^{-1} up **b** 0.192 kg m s^{-1} up **c** 0.591 N up
WE 10.6.3 **a** 32 N **b** 0.36 kg m s^{-1} up
1 **a** 450 kg m s^{-1} east **b** 450 kg m s^{-1} east **c** 129 N east
2 Airbags are designed to increase the duration of the collision, which changes the momentum of a person's head during a car accident. Increasing the duration of the collision decreases the force, which reduces the severity of injury.
3 **a** 1.17 kg m s^{-1} east **b** 1.17 kg m s^{-1} east **c** 11.7 N east
4 3.90 N east
5 **a** 9.0 kg m^{-1} **b** 180 N in the direction of the ball's travel
  c 180 N in the opposite direction to the ball's travel
6 **a** 1200 N **b** 63 N s
7 **a** 1.25 kg m s^{-1} opposite in direction to its initial velocity
  b 1.25 kg m s^{-1} opposite in direction to its initial velocity
  c 1.6 × 10^3 N in the opposite direction to the initial velocity of the arrow
8 **a** The crash helmet is designed so that the stopping time is increased by the collapsing shell during impact. This will reduce the force, as impulse = $F\Delta t = \Delta p$.
  b No. A rigid shell would reduce the stopping time, therefore increasing the force.

Chapter 10 Review

1 No, a force has not pushed the passengers backwards. Since the passengers have inertia, as the train has started moving forwards the passengers' masses resist the change in motion. According to Newton's first law, their bodies are simply maintaining their original state of being motionless until an unbalanced force acts to accelerate them.
2 D **3** 38.3 kg **4** 5.40 m s^{-2} north
5 1.5 m s^{-2} **6** 1.05 m s^{-2} **7** 90 N
8 130 N **9** 5.11 m s^{-1} west **10** 75.0 N north
11 504 kg m s^{-1} west **12** 221 kg m s^{-1} backwards
13 0.558 m s^{-1} backwards or –0.558 m s^{-1}
14 480 kg m s^{-1} N 38.7° E **15** 1.68 N east
16 10 kg m s^{-1} **17** 10 kg m s^{-1} **18** 59 m s^{-1}

Chapter 11 Equilibrium of forces

11.1 Torque

WE 11.1.1 39.5 N anticlockwise **WE 11.1.2** The spanner is long enough. **WE 11.1.3** 17.1 N m **WE 11.1.4** 17.1 N m
1 **a** The magnitude of the torque produced by a given force is proportional to the length of the force arm. By pushing the door at the handle, rather than the middle, the length to the force arm is increased.
  b A crowbar can be used to generate a large torque because the force can be applied at a large distance from the pivot.
2 **a** 200 N m anticlockwise **b** zero
3 0.5 m **4** 18 N **5** 90 N m
6 **a** 4.9 N m **b** 9.8 N m **c** 4.9 N m **7** 45 N m
8 **a** 3.4 × 10^4 N **b** the torque does not change
  c 5.2 × 10^5 N m clockwise about the pivot

11.2 Translational equilibrium

WE 11.2.1 29 000 N up **WE 11.2.2** $F_{T2} = F_{T1}$ = 8487 N
1 D **2** A, B and D
3 **a** 150 N up **b** 40 N west **c** zero **d** 14 N north west
4 3.479 N up **5** 911 N
6 **a** 613 N to the right **b** 613 N to the left **7** 32 N

11.3 Static equilibrium

WE 11.3.1 **a** 1.96 N up **b** 0.30 m
WE 11.3.2 Around reference point Y (the position of the boy), the clockwise torque due to the girl on the plank is equal to the anticlockwise torque due to the pivot point on the plank. So the plank is in rotational equilibrium.
WE 11.3.3 555 N **WE 11.3.4** 36.8 N downwards on the beam
WE 11.3.5 52.0 N

1 The bench will not work successfully. It is not balanced and will topple over because the weight vector from the centre of gravity is outside the base provided by the two supports. To improve the design the student needs to move the supports apart so the centre of gravity is between the two supports.
2 D **3** C **4** 0.750 m
5 $F_L = 1666$ N, $F_R = 3234$ N
6 **a** downwards **b** 9800 N **c** 980 N **7** 0.889 m
8 **a** $F_h = 400$ N $F_v = 693$ N **b** 340 N

Chapter 11 Review

1 B **2** 5.00 N m **3** 0.289 m **4** D
5 15000 N m **6** 1.34 m **7** C **8** 38.6 N m
9 25.3 N m **10** A **11** **a** 0.750 m **b** 9.375 N m
12 0.595 N m **13** 56.5 N m **14** 5.25 kg
15 2116 N **16** 0.765 kg
17 cable A: 1132 N, cable B: −566 N (566 N opposing A)
18 C **19** B **20** A
21 **a** 147 N m anticlockwise **b** 196 N m clockwise
c 49 N m clockwise
22 75 N m clockwise **23** 3.60 N m **24** 74.8 N m
25 51.4 N m **26** 847.7 N **27** 109 N
28 right-hand cable: 122.5 N, left-hand cable: 140.5 N
29 15.5 cm
30 **a** **i** $F_A = 980$ N down, $F_B = 4.9$ kN up, $F_C = 4.9$ up, $F_D = 980$ down
ii $F_A = 1.5$ kN down, $F_B = 5.8$ kN up, $F_C = 5.8$ kN up, $F_D = 1.5$ kN down
b As the woman walks from A to B, the force acting in pillar A decreases and the force acting in B increases. When the woman passes point B and continues on to point P, the forces in both A and B increase in order to produce a greater torque to counterbalance the increase in torque as she moves to point P.

Chapter 12 Energy, work and power

12.1 Work

WE 12.1.1 250 J **WE 12.1.2** 1150 J **WE 12.1.3** 0.05 J
1 10000 J
2 The person exerts a force on the wall but the wall has no displacement (s = 0), so no work is done.
3 1.5 J **4** 18 N **5** 310 J
6 The equation $W = Fs$ applies to situations where the applied force is constant. Since a spring obeys Hooke's law, the force required to compress a spring is not constant.
7 Since the box does not move, no work is done.
8 3.6 J **9** $W_A = 3.0$ J $W_B = 2.4$ J $W_C = 1.0$ J
10 **a** 8 J **b** 4.5 J
c As the basketball bounces, some energy is lost as heat and sound so the work when the ball rebounds is less than the work done when the ball compresses.

12.2 Mechanical energy

WE 12.2.1 78 J **WE 12.2.2** **a** −840 kJ **b** 21000 N
WE 12.2.3 114 km h^{-1} **WE 12.2.4** 15 J
WE 12.2.5 32 J **WE 12.2.6** 8.30 J
1 56779 J or 57 kJ **2** 370000 J **3** 40 km h^{-1}
4 by a factor of 2 **5** **a** 4.6 J **b** 2.3 J
6 2360 kJ **7** 0.304 m **8** 0.78 J
9 Spring A: 1.25×10^{-6} J Spring B: 3.74×10^{-6} J
Spring C: 1.00×10^{-5} J **10** Spring A is stiffer than spring B.

12.3 Using energy: Power and efficiency

WE 12.3.1 50 J **WE 12.3.2** 3.8 m s^{-1} **WE 12.3.3** 76 m s^{-1}
WE 12.3.4 4.19 kJ **WE 12.3.5** 700 W **WE 12.3.6** 26400 W
1 **a** 26 460 J (26000 J to two significant figures) **b** 17 640 J
2 **a** 17 m s^{-1} **b** 14 m s^{-1} **3** 1.5 m
4 **a** 336 J **b** 29.0 m s^{-1} **5** 600 J **6** 1.2 m
7 113 115 W or 113 kW **8** 80 kW **9** 1800 N
10 5.1 m s^{-1}

Chapter 12 Review

1 160000 J **2** 31 J **3** 58800 J **4** 1200 J
5 8.2 m **6** 73.25 N **7** 140 J **8** 11 m s^{-1}
9 by a factor of 2^2 or 4 **10** 340 J
11 $k_A = 4000$ N m^{-1} $k_B = 1000$ N m^{-1} $k_C = 500$ N m^{-1}
12 14.4 m s^{-1} **13** **a** 2.2 J **b** 2.2 J **c** 1.7 m s^{-1}
14 196000 W ≈ 200 kW **15** 35 kW **16** 1500 N
17 **a** 1920 J **b** 96 N **18** 19.2 J **19** 1562.5 J

Unit 2 Area of Study 1 Review

1 C **2** D **3** B **4** C **5** C
6 B **7** B **8** A **9** D **10** A
11 D **12** A **13** C
14 **a** $t = 10.2$ s **b** 510 m **c** 9.8 m s^{-2} downwards
15 13.9 km on a bearing of 21° **16** 0.55 m s^{-1}
17 The weight will travel three times as far during the second second as during the first.
18 **a** 10 m **b** 5.0 s **c** 60 m
19 **a** 5 s **b** 4.08 s **c** 29.9ms^{-1} downwards
20 **a** For the first 30 s, the cyclist travels 150 m east at constant speed, then he accelerates for the next 10 s travelling a further distance of 150 m. He then travels at a higher constant speed for the next 10 s, travelling a further distance of 200 m.
b 5 m s^{-1} east **c** 20 m s^{-1} east **d** 12.5 m s^{-1} east
e 1.5 m s^{-2} east **f** 10 m s^{-1}
21 **a** B, C, D and E (B is a negative acceleration) **b** A
22 **a** C **b** D **c** A
23 **a** 0.10 m s^{-1} **b** 0.30 m s^{-1} **c** 0.50 m s^{-1}
24 **a** 0.10 m s^{-1} **b** 0.30 m s^{-1} **c** 0.50 m s^{-1}
25 The marble is moving with constant acceleration.
26 **a** F_{ME} is the gravitational force on the man by the Earth and $F_N = F_{MS}$, the normal reaction force exerted on the man by the surface.
b $F_{ME} = 980$ N $= F_N$, the forces are equal in magnitude but opposite in direction.
c F_{EM} is the gravitational attraction that the man exerts on the Earth, which is the reaction force to his weight; F_{SM} is the force that the man exerts on the surface.
27 **a** F_{AE} is the gravitational force on A by the Earth and is directed downwards. This is balanced by an equal normal reaction force F_{AB} directed upwards. Both forces are 100 N.
b F_{BE} is the weight of the block B, 100 N directed downwards. F_{BA} is the force exerted by A, also 100 N downwards. The normal reaction force F_{BC} balances both and is hence 200 N upwards.
c F_{CB} is 200 N downwards (effectively caused by the weight of A and B is the reaction pair to F_{BC}), $F_{CE} = 100$ N downwards and the normal reaction force $F_{CT} = 300$ N is exerted upwards by the table.
d The force F_{BE} is still exerted on block B, but $F_{BC} = 0$. Both A and B fall and so the contact forces between A and B also go to zero. Each block only experiences its own weight force and accelerates under gravity.
28 **a** 0.375 m s^{-2} west **b** 375 N west
c 575 N west **d** 575 N east
29 70.6 km h^{-1}
30 **a** 600 N **b** $F = 800$ N **c** 4.0 m s^{-2} up the incline
31 **a** 3.3 m s^{-2} up for the 10 kg mass and 3.3 m s^{-2} down for the 20 kg mass. **b** $T = 1.3 \times 10^2$ N
32 **a** 300 N **b** 1.0 kJ **c** 3.0 kJ **d** 3.0 kJ **e** 17.3 m s^{-1}
33 2.4 m s^{-1} **34** 0.63 J
35 Energy is converted into sound and heat when the mass collides with the pencil.
36 Forces can be equal in magnitude and opposite in direction hence causing no translation because there is no net force. If they do not act along the same line however, there can be a net torque, hence the object is not in rotational equilibrium.
37 450 Nm
38 Thomas should sit 4.6 m from the pivot opposite his siblings.
39 $F = 400$ N

40 **a** 42° **b** 8.9×10^3 J

c There is no displacement in the direction of the vertical component, so no work is done.

d 8.9×10^2 W

41 $F_{\text{column}} = 3307$ N, $F_{\text{ground}} = 367$ N

42 **a** 1.6 MJ **b** 1.6×10^4 N **c** 16.4 kN **d** 1.64 MJ

e 328 kW **f** 40 kJ **g** 97.6%

43 **a** 1.4×10^2 J **b** 80 J **c** 6.4 m s^{-1} **d** 140 J

44 **a** 40 J **b** 250 **c** 5.0 cm

d The elastic potential energy stored in the spring is transferred back to the trolley as kinetic energy when the spring starts to regain its original shape.

e The spring is elastic. This means it can retain its original shape after the compression force has been removed.

45 **a** 1.1×10^4 N m^{-1} **b** 81%

c just before striking the trampoline on first descent

46 Gravitational potential energy ⇒ kinetic energy ⇒ elastic potential energy of trampoline ⇒ kinetic energy as child rebounds losing contact with trampoline (with some loss to heat and sound) ⇒ gravitational potential energy

47 **a** the section from 2.0 cm to 5.0 cm **b** $v = 7.1$ m s^{-1}

c 3.0 J

d The kinetic energy has been converted into heat and sound.

e 100 N

48 The distance of the race = area under the v–t graph from $t = 0$ s to $t = 11$ is $s = 100$ m.

49 **a** 9.1 m s^{-1} **b** 8.0 m s^{-1}

50 **a** A, B, C **b** D **c** E, F, G **d** A, E **e** C, G

51 **a** 160 m **b** 2.0 m s^{-1}

52 **a** 1.2 m s^{-1} north **b** 5.4 m s^{-1} north **c** 3.0 m s^{-1} south

53 **a** 0.69 m s^{-2} **b** 5.7×10^2 N **c** 2.5×10^2 m

d Kinetic energy is converted into heat.

54 The swimmer pulls against the water with her arms, exerting a force on the water. The reaction force (Newton's third law) of the water on her arms is what propels her forwards. If this reaction force is greater than the sum of the drag forces on her, she will accelerate according to Newton's second law. If there is no net force she will travel at constant speed according to Newton's first law.

55 Earth: 1.96×10^5 N Moon: 3.2×10^4 N

56 **a** The chair will obtain an initial velocity from the push, and then quickly slow to a stop due to sliding friction.

b Castors will roll, resulting in less friction, hence the chair will travel further before coming to a stop.

c In both cases the force of the initial push causes the chair to accelerate while being pushed, which is an application of Newton's first and second laws. While the person pushes on the chair (action) the chair will push back on the person (reaction) with an equal and opposite force, an example of Newton's third law.

57 The steady force applied by the engine is equal and opposite to the combined resistance forces such as air resistance, friction between the wheels and track etc. The net resultant force on the carriages is zero, and according to Newton's first and second laws, constant velocity is the result.

58 **a** 85 N m^{-1} **b** 43 J

59 **a** 645 kg m s^{-1}

b Momentum is always conserved. The motion of the pole is minute because it is anchored in the ground and very heavy.

c As his skull comes to rest against the pole, according to Newton's first law, his brain would continue in its motion at 7.5 m s^{-1} until it collides with the skull, incurring potential damage in the collision.

60 **a** 2.2 m s^{-1}

b E_k before $= \frac{1}{2}mv^2 = 1.8 \times 10^5$ J; E_k after $= 1.3 \times 10^5$ J, hence kinetic energy is not conserved.

c Total energy is always conserved but kinetic energy is only conserved for perfectly elastic collisions. In this case there is considerable loss to heat and sound in the collision.

61 **a** 10 kg m s^{-1} **b** 10 kg m s^{-1} **c** 59 m s^{-1}

62 **a** 1.6×10^4 m s^{-2} **b** 8.8×10^6 N **c** 4.4×10^5 kg m s^{-1}

d 4.1 m s^{-1} **e** 8.8×10^6 N

f Work $= 1.8 \times 10^8$ J; E_k of shell $= 1.8 \times 10^8$ J

This obviously represents an ideal situation; realistically there would be significant losses.

Chapter 13 Stars

13.1 Astronomical measurements

WE 13.1.1 **a** 4.6×10^{14} Hz **b** 2.2×10^{-15} s

WE 13.1.2 1 m

WE 13.1.3 **a** 1.3 pc **b** 3.9×10^{16} m

WE 13.1.4 100 Mpc

13.1 Review

1 8.6×10^{14} Hz **2** 3.75×10^{16} Hz **3** 2.6×10^{-17} s

4 The shift toward the red end of the spectrum in light from distant galaxies.

5 3600 **6** 2.27 pc or 6.81×10^{16} m **7** 0.641 pc

8 61-Cygni is much closer to Earth than 10 pc.

9 Betelgeuse must be much further away than 10 pc.

10 C **11** 1000 pc or 1 kpc

13.2 Classifying stars

WE 13.2.1 the blue star

13.2 Review

1 the luminosity of a star (which is derived from the absolute magnitude) against the spectral type of stars (from which the temperature of the star is derived)

2 Along the main sequence, luminosity increases with the surface temperature.

3 The continuous spectrum provides information about the surface temperature of the star. The absorption spectrum gives information about the elements present.

4 By determining what type of star it is from its spectrum, it can then be placed on the diagram according to its spectral type (OBAFGKM). This enables the luminosity (or absolute magnitude) to be found. The distance can be determined from a comparison of the luminosity with the apparent magnitude.

5 a red giant

6 Betelgeuse; Vega; Sirius A; the Sun; Sirius B; Proxima Centauri

7 The size of a star cannot be seen even using the best telescopes and so must be determined indirectly. Once the temperature is known, the amount of power given off per square metre can be calculated. This is compared with the total luminosity of the star to find the total surface area and hence the radius.

8 C **9** D **10** A

13.3 The life and death of stars

WE 13.3.1 5.6×10^8 kg s^{-1} **WE 13.3.2** 89 km

13.3 Review

1 Nuclear fusion reactions in the Sun involve fusing hydrogen *nuclei* to produce *helium* nuclei. In contrast, hydrogen burning in oxygen involves only the electrons in the outer shell of the atoms. The energy involved in nuclear reactions is about 100 million times greater than chemical reactions.

2 C **3** 5.4×10^{26} L

4 protostar → main sequence star → red giant → white dwarf → black dwarf

5 planetary nebulae

6 It is most likely to initially expand into a red giant before collapsing and becoming a white dwarf.

7 the Schwarzschild radius

8 74 km **9** 2.8 m

10 **a** red sequence **b** blue cloud **c** green valley

Chapter 13 Review

1 1.6×10^{11} Hz **2** 6.2×10^{-12} s **3** B

4 Canopus is many times further away from Earth than Sirius, which means that when viewed from Earth is will not appear as bright.

5 D

6 **a** continuous **b** emission **c** continuous
d absorption **e** emission

7 spectral class, surface temperature and chemical composition

8 Their spectra show just the same lines as our Sun and these lines correspond to the 98 known elements in our periodic table.

9 Barnard's Star, Tau Ceti, Vega, Aldebaran, Rigel

10 **a** white dwarfs **b** main sequence
c supergiant stars **d** giant stars

11 Like all stars Rigel would have started from a dust and gas cloud collapsing to form a protostar.
As Rigel starts to convert silicon to iron as the main fusion process, less energy will be produced than needed for the fusion process. It will begin to collapse.
As a giant star, the core of Rigel can then be expected to heat to billions of degrees in a fraction of second. An explosive supernova results.
The final stage will be either a neutron star or a black hole.

12 into the middle (approximately) of the main sequence

13 The Sun is close to the centre of the H–R diagram; that is, in terms of the overall range, it is of average temperature and average brightness. However, most stars are actually cooler and duller than the Sun.

14 5.9×10^4 m **15** a singularity

16 The observer will see light near the black hole slow down, theoretically to the point where an object falling past the event horizon of a black hole would appear frozen at the black hole's edge.

17 A, B and C **18** 3.6×10^{26} J **19** 8.9 mm

20 This is approximately 34 $M_{\odot}$

Chapter 14 Forces in the human body

14.1 Forces acting on the body

WE 14.1.1 at the centre of mass, 0.94 m from the origin

14.1 Review

1 **a** B **b** B **c** A **d** E

2 B and D **3** A **4** B **5** C

6 roughly above the navel

7 lower your centre of mass and broaden your base of support by standing with your feet slightly further apart

8 The bone bends as a result of the applied force in such a way that the side closest to the applied force is under compression and the side opposite the force is under tension. Bone is weaker under tension than compression, so that is where the fracture starts.

9 A heavy impact from the side would cause bending such that the side of the impact would be under compression and the opposite side would be under tension. Since this material is weaker under compression, it is likely to fail on the side of the impact first.

10 1.1×10^2 cm (to one decimal place)

11 The centre of mass of each of the parts—head, torso, arms and legs—would lie within that body part.
Since there is considerable mass from legs and arms below the torso, this moves the centre of mass lower.
The overall centre of mass is outside her body (in the same way that the centre of mass of a ring donut is in the middle of the hole). Her weight can be considered to act downwards from the centre of mass. The weight creates a torque causing her to fall forwards. Her feet cannot provide a balancing force because her weight acts beyond the base of support.

14.2 Forces cause rotation

WE 14.2.1 16 N m **WE 14.2.2** 11 N m

14.2 Review

1 A and C

2 Head: F_2 is the effort and F_1 is the load.
Arm: F_4 is the effort and F_3 is the load.
Leg: F_5 is the effort and F_6 is the load.

3 Skull–neck joint: class 1 lever, speed multiplier
Elbow joint: class 3 lever, speed multiplier
Toe joint: class 2 lever, force multiplier

4 $\tau = d_2 F \sin\theta$ **5** B and C

6 **a** 45 N m **b** 890 N **7** 27 N

8 T_Y is the only component of force that provides an anticlockwise torque about the sacrum. In order to lift an object, it must overcome the clockwise torque created by the weight of the head, torso and load. Since the tension, T, is at such as small angle from the horizontal (12°), T_X will be significantly larger than T_Y (by a factor of $\frac{1}{\tan 12} = 4.7$). R_X must be equal in magnitude to T_X to maintain a balance of horizontal forces. Therefore, T_X and R_X put the back under considerable compressive stress, which can cause injuries such as slipped discs.

9 F_B 5.5×10^2 N

14.3 Tissue under load: Stress and strain

WE 14.3.1 1×10^6 N m^{-2} **WE 14.3.2** 0.080 **WE 14.3.3** 29 cm

14.3 Review

1 C **2** 2σ

3 **a** 3.9×10^7 N m^{-2} **b** 6.3×10^3 N

4 2.0×10^2 mm **5** 3.3×10^{-4} **6** 8.7×10^5 N m^{-2}

7 D

8 $\sigma = 3.4$ MPa $\varepsilon = 0.04$ **9** A, B and E

10 $\sigma = 7.2 \times 10^7$ N m^{-2} $\varepsilon = 0.05$

14.4 Properties of human tissues

WE 14.4.1 0.18 MPA **WE 14.4.2** 2.5 MJ m^{-3} **WE 14.4.3** 5.2×10^2 J

14.4 Review

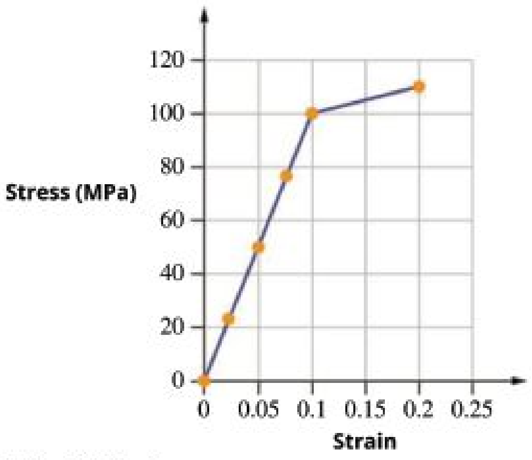
1 A

2 **a** bone, hair, resilin
b bone, hair, resilin
c hair, resilin, bone
d Hair has the greatest plasticity, whereas both bone and resilin are brittle with no plastic region.

3 C, E, F and G

4

Stress–strain curve for tendon



5 1.0×10^9 N m^{-2}

6 **a** 4.0×10^8 Pa (or N m^{-2})
b 2.0×10^{-3}
c 2.0×10^{11} N m^{-2}
d Young's modulus
e 6.0×10^8 Pa (or N m^{-2})

7 **a** 5.0 mm **b** 2.0 mm

8 **a** 1.6 mm **b** zero

9 **a** ductile **b** 3.0 mm
c No, since the material does not behave elastically under this tensile stress.

10 **a** 4.0×10^5 J m^{-3} **b** 16 J
c It is in the elastic region of the graph, so will resume its original length and no energy will be transformed into heat.

11 **a** 1.35×10^6 J m^{-3}
  b 54 J
  c It is past the elastic limit, so the alloy will be permanently deformed and will heat up as it is stretched.

12 **a** The alloy will break.
  b 1.9×10^6 J m^{-3}
  c 76 J

14.5 The future: Materials for use in prosthetics

1 **a** Prostheses should not corrode or be brittle.
  b Prostheses should be able to sustain high stress without failure.
  c Prostheses should give under load to some extent to avoid jarring.
  d Prostheses should not deform excessively under stress.
  e Prostheses should be lightweight in order not to overload muscles.
  f Prostheses should absorb energy and return it in a similar way to the structure being replaced (e.g. tendons).

2 **a** During systole, neither collagen nor elastin is very stiff. During diastole, both materials become stiffer, but collagen becomes stiffer than elastin.
  b elastin
  c collagen

3 E **4** D

5 Internal prostheses are more difficult. Both types need to mimic the properties of the biological tissues they replace but internal prostheses must not shed particles into the bloodstream, must not be toxic, must not corrode, and need to be utterly reliable because they are harder to replace.

Chapter 14 Review

1 B **2** A and B

3 Tendons are essentially fibres that buckle under compression. The collagen that they are made from straightens out and then becomes stiff under tension. When stretched, tendons can absorb large amounts of energy without failing.

4 toughness

5 brittle materials

6 B and D

7 A and B

8 A, B and D

9 A, B, C, E and G

10 **a** compressive stress; **b** bending; **c** shear stress; **d** tensile stress.

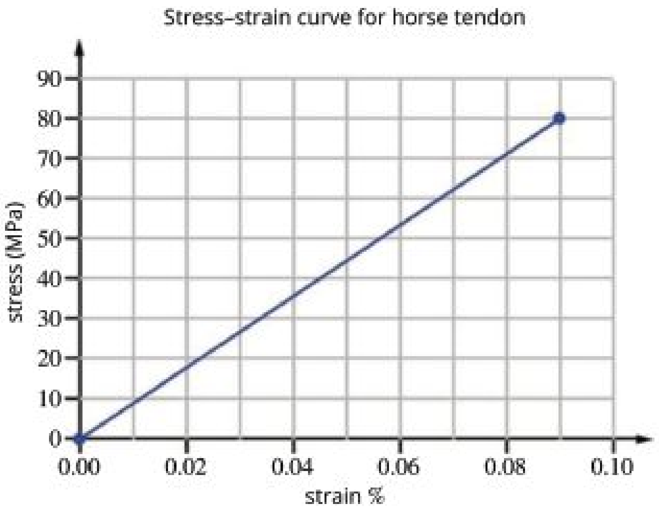
11 At A the fibres fail and rupture.
At B the fibres are stretching in proportion to the applied stress and becoming stiff.
At C the fibres are uncrimping and experiencing a large strain with little applied stress.

12 **a** Longitudinal direction—the longitudinal curve demonstrates the greatest stress value under tension.
  b longitudinal **c** transverse **d** longitudinal
  e longitudinal **f** longitudinal **g** longitudinal

13 'Anisotropic' means that a substance is not uniform in all directions. It is clear that bone has significantly superior material properties for load-bearing in the longitudinal direction. This is the direction in which bone is routinely stressed, and so these superior properties permit bones to perform their function of load-bearing. Bones are rarely stressed in the transverse direction.
When a bone experiences as transverse force it tends to bend. It is then under tension on one side and under compression on the other. Bone is weaker under tension in the transverse direction as may be seen from the graph. The bone is likely to fracture on the convex side of the bone.

14 **a** A **b** B and C **c** B and C **d** A

15 80×10^6 N m^{-2} or 80 MPa



16 3.6×10^6 J m^{-3}

17 Pack fits snugly against the back so that the centre of mass of the body is not moved too far backwards, leading to instability. Chest straps should be fastened to keep the pack close to the body, again to keep the centre of mass closer. The hip belt is firmly fastened so that most of the load is transferred from the pack via the internal frame directly onto the pelvis. This prevents compression of the spinal column, which would occur if the weight were carried on the shoulders. Heavier objects are packed closer to the back and lighter objects further from the body, keeping the centre of mass of the pack close to the body.

18 **a** The carbon fibres themselves are very strong under tension when stressed in the longitudinal direction (i.e. along the length of the fibre).
  b The matrix, or binding material, which is usually a polymer or epoxy resin. Carbon fibres themselves do not have rigidity and therefore do not offer compressive strength.
  c By using a composite laminate, which means layers of the carbon fibre composite are stacked on top of each other in different directions.

19 5.5×10^4 N

20 **a** 2.5 mm
  b Tension, since the tensile strength for bone is lower than the compressive strength for bone.
  c As we age, calcium loss causes the cross-sectional area of our bones to decrease. For any given load, therefore, the stress that the bones will experience will be greater and they will be more likely to fracture

21 1.2×10^7 Pa (or N m^{-2})

22 **a** C, D **b** A **c** D **d** B **e** A

23 In elastic deformation, all the strain energy is returned by a material as it regains its initial dimensions after the strain is removed. For plastic deformation, some of the strain energy will be transformed into heat energy in altering the atomic structure of the material. Consequently, the material in this situation never regains its original dimensions.

24 **a** 9.55×10^7 Pa (or N m^{-2})
  b 2.28×10^4 J m^{-3}
  c 0.143 J

25 No, the rod fails at a point past the elastic limit, where the behaviour of the material is nonlinear. A stress–strain graph is needed.

26 1.5

27 **a** 9.5×10^5 J m^{-3} **b** 6.5×10^5 J m^{-3}

28 Material P, because it can absorb a greater amount of strain energy per unit volume before failing (indicated by area under stress–strain graph).

29 **a** Material P, since it has a greater value for Young's modulus (indicated by gradient).
  b Material P, since it experiences a greater stress value before failing.

Chapter 15 Energy from nuclear power

15.1 Energy from the nucleus

1 D **2** C
3 **a** uranium-238
 b Because fast neutrons are used
 c Breeder reactors use nuclear fuel enriched in fissile material to transform U-238 into fissile P-239 and produce more P-239 than they use U-238.
4 **a** absorption + ^1_0n **b** decay β^- **c** decay β^-
5 **a** More neutrons released will generate more fissions, and hence more energy.
 b Pu-239, because there are more neutrons able to generate more fissions.
6 Flat sheets allow more neutrons to escape, so the proportion of neutrons escaping from each fission reaction is too high to sustain fission.
7 two; neutrons; vary; conserved
8 When the positrons come into contact (almost immediately afterwards) with electrons, they are annihilated, releasing the additional energy. Each of the two positrons (e^+) annihilates with an electron (e^-) to release an additional 2 × 1.02 MeV, accounting for the additional energy making up the total energy released of (24.7 + 2 × 1.02) = 26.74 MeV.
9 Deuterium–tritium (the H^{-2} and H^{-3} isotopes of hydrogen respectively). It was successfully initiated in a research reactor in 1997.
10 Limited fuel source: While tritium does not exist in nature, it can be produced artificially. Lithium is the limiting fuel source in this process and hence limits the overall fuel available for deuterium–tritium fusion reactors.
 Very high temperatures: a temperature of 4×10^7 K is required to overcome the repulsive forces of the like-charged protons in each nucleus to initiate fusion.

15.2 Nuclear energy as a power source

1 **a** coolant: keeps core from overheating
 b fuel rod: location of fission reactions
 c moderator: slows fast neutrons
 d heat exchanger: transfers energy from coolant to produce steam
 e control rod: absorbs neutrons
2 **a** fuel rods: nuclear energy
 b heat exchanger: heat energy
 c generator: kinetic energy
 d power distribution: electrical energy
3 A and D
4 No control rods would mean no neutrons would be absorbed, so the chain reaction that resulted will be uncontrolled.
5 A **6** 1.5 million tonnes
7 an isotope of hydrogen, found in lakes and oceans
8 Few radioactive particles are produced by fusion reactions. Hydrogen isotopes deuterium and tritium (produced in the nuclear reactor from lithium) are the primary fuels. Deuterium is not radioactive nor is the lithium used in 'breeding' tritium. Tritium is radioactive but has an extremely short half-life (12.6 years) when compared with the radioactive waste of fission reactors. It can be produced solely within the reactor itself. Helium, the product, is inert and is also safe.
 Like fission, fusion produces no greenhouse gases or other atmospheric pollutants. Little radiotoxicity would remain once a fusion power station reached the end of its working life and was decommissioned.
9 Over a period of months, the fissile nuclei in the fuel rods become depleted (used up), the number of fissions decreases, and so fewer neutrons are flying around in the core. In order to maintain the chain reaction, the control rods must be gradually withdrawn by raising them.
10 a low percentage of U-235

Chapter 15 Review

1 C
2 **a** not likely to undergo fission
 b U-238 is first transformed into Pu-239, which is fissile.
 c U-238 is highly abundant.
3 B
4 neutrons; conserved; mass defect; energy
5 X = 2
6 Uranium-235 is more fissile with slow neutrons.
7 **a** uranium-238 **b** uranium-235
 c The concentration of the fissile uranium-235 is too low.
8 First, a neutron causes fission to occur in a uranium-235 nucleus, thus releasing 2 or 3 more neutrons. These then go on and induce fission in more uranium-235 nuclei, each resulting in the release of 2 or 3 neutrons, and so on. The chain reaction grows very rapidly and energy is released in each fission reaction.
9 As a result of its shape, a very high proportion of neutrons can escape from the flattened material, and so the chain reaction dies out.
10 **a** chain reaction will die out **b** steady chain reaction
 c chain reaction will grow, potentially uncontrolled
11 The tokamak style of reactor has a torus-shaped core where the plasma of the fusion fuel is located. This fusion material is held away from the reactor wall by magnetic fields. The heat energy produced is carried from the reactor core by lithium and water.
12 $4\,^1_1H \rightarrow \,^4_2He^+ + 2e^+ + 2\nu + 2\gamma + 24.7$ MeV
13 **a** The fission process in the reactor core produces heat. This heat energy is conducted into the coolant that is flowing through the core. The energy is used to produce steam which drives a turbine to generate electricity.
 b The steam is produced by burning coal instead of a nuclear fission reaction.
 c They both use steam to turn a turbine to generate electricity.
14 **a** absorbs neutrons
 b provides shielding to prevent radiation escaping
 c slows the neutrons
 d stops the core from overheating
 e contains enriched uranium
 f removes heat from core
15 2.7×10^3 tonnes
16 fission; coolant; steam
17 Uranium-235 is more fissile when struck by slow (thermal) neutrons. A moderator is used to slow down the neutrons released during fission. Collisions between the moderator material and the neutrons slow the neutrons down. Water, graphite, heavy water and carbon dioxide are used as moderators.
18 Each fission reaction releases much more energy than each fusion reaction, 200 MeV and 25 MeV respectively. The energy is created when the mass of the nucleons decreases slightly in both fission and fusion. So energy is created by the slight loss of mass of nucleons, not the loss of nucleons. The number of nucleons does not change in either process.
 The amount of energy can be determined by using $E = mc^2$. However, the percentage of mass lost is greater for fusion than for fission. Therefore, the energy released per nucleon is greater during fusion than during fission.
19

Advantages of fusion reactors	Disadvantages of fusion reactors
deuterium is virtually unlimited no radioactive waste	fusion requires very high temperatures to initiate lack of naturally occurring tritium potentially limits fuel

20 **a** Intermediate-level waste (ILW) requires shielding. If it has more than 4000 Bq g^{-1} of long-lived (over a 30-year half-life) alpha emitters it is categorised as 'long-lived' and requires more sophisticated handling and disposal.
 Australia does not have any high-level waste. It is spent fuel from the research reactor. This waste would be defined as 'intermediate-level', and the waste which results from

reprocessing it abroad will be returned as long-lived waste for disposal or storage.

b Most commercial reprocessing takes place in France (La Hague). Reprocessing involves dissolving spent nuclear fuel in acid and separating the unused uranium (about 96% of the mass), plutonium (1%) and high-level wastes (3%).

c The following table summarises the current levels of waste in Australia.

Waste type	Waste description	Commonwealth waste inventory (approximate volume)
Low-level waste	• lightly contaminated laboratory items such as paper, plastic and glassware • operational waste from the research reactor • contaminated items from production of radiopharmaceuticals • research reactor decommissioning waste such as graphite, concrete and steel • lightly contaminated soil	total: 4048.28 m^3
Intermediate-level waste	• higher activity operational waste from ANSTO including irradiation cans, ion exchange resins, aluminium end pieces of fuel rods, control arms and general waste from radiopharmaceutical production • concentrates from mineral sands processing (thorium and uranium residues)	total: 551.5 m^3

Chapter 16 Nuclear medicine

16.1 Producing medical radiation

1 Radio waves, microwaves, infrared radiation, visible light, UV radiation, X-rays and gamma rays.

2 ionising, soft X-rays

3 ionising, hard X-rays

4 neutrons

5 **a** ${}^{99m}_{43}\text{Tc}$

b ${}^{99}_{43}\text{Tc}$

6 Soft X-rays and hard X-rays are *ionising*. Soft X-rays have *greater* wavelengths than hard X-rays.

7 Iodine-131 emits high energy β-particles which are very effective at destroying cancer cells.

8 uranium-238

9 Hard X-rays are used for *therapeutic* purposes and soft X-rays are used for *diagnostic* purposes.

10 Kinetic energy (of the electrons) is converted to electromagnetic energy (of the X-rays).

16.2 Measurement of radiation doses

WE 16.2.1 4.0 Sv

16.2 Review

1 Somatic effects arise when ordinary body cells are damaged. Examples include skin rash, hair loss and nausea.

2 Somatic. Genetic damage affects the DNA and is transmitted throughout generations.

3 2 Gy

4 They are all equally damaging as they are already in the units of dose equivalent, sieverts (Sv).

5 **a** 2.0×10^{-4} Sv **b** 1.6×10^{-2} J

6 Here 20 μGy of alpha radiation has the highest DE of 400 μSv.

7 **a** a little over 2 days **b** 879 mSv

8 90 hours

9 Ovaries have the highest *W* value of 0.20, therefore these organs are the most radiosensitive.

10 14.75 mSv

16.3 Radiation in diagnosis and treatment of human disease

1 An X-ray beam is sent into the tumour for the appropriate *time*. The beam is *rotated* so that healthy cells do not receive large doses of ionising radiation.

2 X-rays

3 Any two of:

- have a short half-life (hours or days) that is appropriate for the time taken for the diagnostic procedure. Radioactive materials are considered to be relatively safe after around 10 half-lives have passed.
- emit only radiation of an energy that can be detected by a camera
- not emit alpha or beta radiation because these particles would be trapped in the patient's tissues and they would not be detected externally
- be available in the highest possible activity but not be toxic to the patient or react with drugs used at the same time.

4 MRI

5 One of a group of drugs which has a radioactive tracer attached to it.

6 *Positron* emitters. These particles interact with electrons from the body cells and create gamma rays which are detected externally by a *gamma* camera.

7 The small amount of harm caused by the radiation is outweighed by the benefits of the diagnosis.

8 Phosphorus-32

9 MRI scans are similar to CT scans because *slices* of the body parts are photographed. An MRI uses *magnetism (non-ionising)* to produce its images whereas CT scans use X-rays (ionising).

10 There is *no* radiation dose.

Chapter 16 Review

1

Electromagnetic radiation	Particle radiation
UV X-rays gamma	alpha beta cosmic rays

2 Soft X-rays have *less* energy and penetrating power than hard X-rays. Soft X-rays and hard X-rays move at the same speed.

3 ${}^{31}_{15}\text{P}$

4 ionising; electromagnetic radiation

5 ionising radiation

6 The dose of radiation is not being delivered in one treatment session. In order to deliver this dose of radiation, the patient would receive the dose over multiple treatment sessions.

7 1 Gy of alpha. Since alpha has a quality factor of 20, 1 Gy of alpha radiation is 20 times more damaging than 1Gy of the beta and gamma.

8 They are all equally damaging as they are already in the units of dose equivalent, sieverts (Sv).

9 300 mSv

10 **a** 380 J (to two significant figures)

b 5.0 Sv (to two significant figures)

11 **a** 4 μSv per X-ray **b** 4 times

12 **a** 10 μSv **b** 0.50 μSv **c** 0.50 μSv

13 350 μSv

14 technetium; it is produced in a nuclear reactor

15 non-ionising radiation; magnetic fields and radio waves

16 Tracers need to able to pass through body tissue and be detected externally. Gamma rays are used as diagnostic tracers because they can pass through the body whereas alpha and beta do not easily pass through the body.

17 a *radioactive* tracer that emits *positrons*

18 Positrons interact with *electrons* in the body tissue and create *gamma rays* that are detected by a *gamma* camera.

19 SPECT (single photon emission computed tomography) is used for *diagnostic* purposes. The patient is injected with a radioactive *gamma emitter* and a *gamma* camera is used to obtain an image.

Chapter 17 Particle accelerators

17.1 Synchrotrons

1 charged particles
2 They are passed through an electric field.
3 linac, booster ring, storage ring, beamlines
4 very close to the speed of light
5 In the storage ring electrons orbit for hours at a time at speeds near that of light, before being channelled along the beamlines for experimentation.
6 in the booster ring
7 in the linac
8 High-intensity, high-energy sources allow X-ray diffraction techniques to be completed over considerably shorter times than with traditional X-ray sources.
9 Synchrotrons produce a continuous spectrum of radiation, and particular wavelengths within that spectrum are highly selectable or tunable.
10

X-ray tube	Synchrotron
single burst	produced over hours
relatively divergent	highly collimated
standard intensity	many times brighter

17.2 Colliders and particle physics

1 from the outward spiralling circular path of the particles
2 Higher energies are achieved by synchronising the accelerating voltage with the particle velocity.
3 It essentially gives mass to all elementary particles.
4 dark matter and antimatter
5 hydrogen atom source, electrons stripped away, linac, proton synchrotron booster, super proton synchrotron, storage ring, detectors
6 The LHCb is able to detect b quarks by having a series of sub-detectors at points largely forward of the collision point in order to focus on particles thrown forward of the collision.
7 2.2×10^{-6} J
8 ALICE seeks to investigate quark–gluon plasma by studying the extremely high temperatures generated by heavy-ion collisions.
9 It receives protons from the linear accelerator. It boosts the initial energy level from the linac of 50 MeV to 1.4 GeV, after which they are fed into the Proton Synchrotron.
10 The LHCb has been designed to look for the bottom or beauty quark. Every collision event forms a broad range of quarks for detection. Since quarks decay quickly, the LHCb has movable tracking detectors forward of the collision to focus on particles thrown forward.

17.3 The importance of accelerator technology in society

1 around 1 MB or 1 million bytes of data
2 They are used for initial storage and analysis of data from the LHC before data is distributed through the Worldwide LHC Computing Grid.
3 Synchrotron light is ideal for X-ray diffraction studies needing a high degree of definition since it is high energy, highly collimated and can be produced over several hours.
4 It is the process of printing a pattern onto another surface.
5 It is high intensity, highly collimated and can be produced continuously for several hours.
6 a heated cathode
7 1.6×10^{-11} J
8 the world's first anti-influenza drug
9 X-ray diffraction
10 a material that will conduct electricity only under particular conditions, placing its properties somewhere between those of a conductor and an insulator

Chapter 17 Review

1 Electrons are 'boiled' off a heated wire element acting as a cathode.
2 linear accelerator
3 A beamline is typically a stainless steel tube of 15–35 m in length along which synchrotron light travels from the storage ring, where it is produced, to its target for experimental work.
4 X-ray diffraction
5 in the electron gun
6 Synchrotrons produce synchrotron light with frequencies ranging from the infrared region through to highest frequency X-rays.
7 3.2×10^{-14} J
8 Hadrons. The other particles listed are fundamental particles.
9 An accelerator capable of generating the 2 TeV accelerated particles required wasn't available before the early nineties.
10 ATLAS: Head-on proton–proton collisions
ALICE: Heavy-ion detector
CMS: General-purpose detector
LHCb: Beauty-quark detector
11 297×10^{6} m s^{-1}
12 A collider features two storage rings, each accelerating particles in opposite directions. Particles in each ring are accelerated to 7 TeV. Colliding the two particles combines the individual energies to give a total of 14 TeV.
13 Any two from the following:
- It is unable to explain dark matter.
- It is unable to explain the absence of antimatter in the Universe today.
- It is unable to provide a completely unified model for all fundamental forces; interactions with gravity are unexplained.

14 nearly 109000
15 Ions are implanted in semiconductors in very specific locations and concentrations allowing the production of higher-density semiconductors, which can mean smaller chips.
16 particle accelerators
17 heated cathode, filter magnet, accelerator column, mass-analysis magnet, scanner
18 Particles are colliding within the LHC approximately 600 million times per second and each collision event generates about 1 MB (1 million bytes) of data. A powerful computer centre is required to store and analyse this massive amount of data.
19 Tier 2 centres are largely universities.
20 Around three years after the initial collision event and after the data has been accessible through the Worldwide LHC Computing Grid, the collision data is made available to anyone including teachers and students.

Chapter 18 Sport

18.1 Collisions

WE 18.1.1 0.83

18.1 Review

1 B
2 The ball bounces too high for use in a tournament. Its value of ω is 0.80 which is above the given range of 0.73–0.76.
3 0.56 m **4** 11 m s^{-1}
5 **a** **i** 0.8 kg m s^{-1} **ii** 5 m s^{-1} **iii** 2 J **iv** 2 J
b elastic
6 The red ball is stationary and the white one moves at 4 ms^{-1}.
7 8 m s^{-1}
8 **a** 0.3 m s^{-1} **b** The collision is inelastic.
9 **a** 9.8 kg m s^{-1}
b 7.5 m s^{-1}
c The collision is inelastic.

18.2 Sliding and rolling

WE 18.2.1 –8.5 N **WE 18.2.2** 6.1 m s^{-1}
WE 18.2.3 **a** 2.27 **b** 324°
WE 18.2.4 314 rad s^{-1} **WE 18.2.5** 15 m s^{-1}

18.2 Review

1 C **2** C **3** 38 m s^{-1}
4 2.0 m **5** 0.46 **6** C

7

Angle (in degrees)	Angle (in radians)
57°	$57 \times \frac{\pi}{180} \approx 0.995$
125°	$125 \times \frac{\pi}{180} \approx 2.18$
280°	$280 \times \frac{\pi}{180} \approx 4.89$
450°	$450 \times \frac{\pi}{180} = 7.85$
$0.50 \times \frac{180}{\pi} \approx 28.6°$	0.50
$1.25 \times \frac{180}{\pi} \approx 71.6°$	1.25
$4.80 \times \frac{180}{\pi} \approx 275°$	4.80
$7.12 \times \frac{180}{\pi} \approx 408°$	7.12

8 Since $w = \frac{\theta}{t}n$

Angle, θ (rad)	Time, t (s)	Angular speed ω (rad s^{-1})
3.14	0.01	$\omega = \frac{\theta}{t} = \frac{3.14}{0.01} = 314$ rad s^{-1}
48	0.3	$\omega = \frac{48}{0.3} = 160$ rad s^{-1}
630	21	$\omega = \frac{630}{21} = 30$ rad s^{-1}
$\theta = \omega \times t = 25 \times 2.4 = 60$ rad	2.4	25
$\theta = 54 \times 0.34 = 18$ rad	0.34	54
$\theta = 8.8 \times 0.1 = 0.88$ rad	0.1	8.8
55	$t = \frac{\theta}{\omega} = \frac{55}{6.0} = 9.2$ s	6.0
8.4	$t = \frac{8.4}{38} = 0.22$ s	38

9 9.30 m s^{-1}

10 Linear velocity = 7.32 m s^{-1} Angular speed = 67.8 rad s^{-1}

18.3 Hitting, kicking or throwing

WE 18.3.1 1.7 s

18.3 Review

1 D

2 Gravitational potential energy is converted to kinetic energy and back throughout the golf swing. The gravitational potential energy is maximum because the head of the golf club is at its highest point. Kinetic energy is zero because the golf club will have stopped moving.

3 D **4** 0.33 J **5** 2.7 s

6 248 mm **7** A

8 Treat the footballer's leg as a simple pendulum. Mechanical energy is conserved in a simple pendulum. By raising her foot, the footballer has given it gravitational potential energy. When the foot swings down to kick the ball, its gravitational potential energy will transform into kinetic energy. The more kinetic energy her foot has, the more kinetic energy will be transferred to the ball and the faster the ball will go.

9 A pendulum consists of a bob, suspended from a fixed point. A double pendulum consists of one pendulum suspended from another pendulum so the upper leg or thigh acts as the upper pendulum; the lower leg or calf acts as the lower pendulum.

10 A double pendulum is governed by the laws of physics but sensitive to the initial position of each bob. The results of its swing will vary depending on the exact starting position of each of the bobs. Small changes in the initial position of one bob will mean that the double pendulum system will swing a very different way.

18.4 The flight of a ball

WE 18.4.1 **a** 0.55 s **b** 11 m
c 20.7 m s^{-1} at 15° down from the horizontal.
WE 18.4.2 **a** 1.4 s **b** 11.20 m (up)

18.4 Review

1 It does not change throughout the shot. It is constant.
2 at the top of its flight
3 C
4 vertical component: 20 m s^{-1}; horizontal component: 30 m s^{-1}
5 5.3 m s^{-1} (up)
6 **a** 4.1 s **b** 123 m
7 5.42 m s^{-1} **8** 54 m

18.5 Air resistance

WE 18.5.1 2.0 N **WE 18.5.2** 23 m s^{-1}

18.5 Review

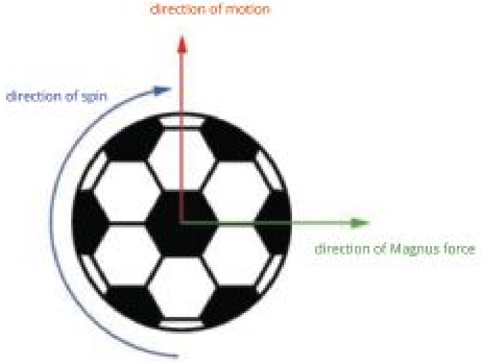
1 D **2** B
3 glycerine, water, air, vacuum
4

Ball type	Diameter (mm)	Radius (m)	Cross-sectional area (m^2)
netball	**226 mm**	0.113 m	0.040 m^2
cricket	72 mm	**0.036**	0.0041 m^2
tennis	68 mm	0.034 mm	**0.0036**

5 0.89 N **6** 4.0 m s^{-1} **7** 30 m s^{-1} **8** 0.27

9 The cricket ball has a much higher mass than the tennis ball; therefore, its weight will be greater. Also, since the cricket ball has a lower drag coefficient, it will experience less drag force. Both of these factors increase the terminal velocity of the cricket ball compared to that of the tennis ball.

10



Chapter 18 Review

1 B **2** 7 m s^{-1}
3 **a** 8.1 J **b** 8.1 J **c** elastic
4 **a** yes
  b inelastic
  c 2.0 m s^{-1}
5 $a = -\mu g$ **6** 0.48
7 –4.1 N $F_f = \mu F_N$ $F_f = 0.52 \times -4.1 = -2.1$ N
8 **a** 22.4 m
  b linear velocity = 5.45 m s^{-1}
  angular speed = 97.3 rad s^{-1}
9 **a** 3.73 m s^{-1} **b** 62%
10 gravitational; kinetic; double pendulum
11 17.8 J **12** 5.2 m s^{-1}
13 **a** 1.89 s **b** 75.6 m **c** 44 m s^{-1}
14 **a** 0 m s^{-1} **b** 7.8 m s^{-1} down **c** 3.1 m **d** 9.3 m s^{-1} **e** 57°

15 **a** 3.71 s **b** 31.4 m

16 **a** 4.0 m s^{-1} **b** 6.9 m s^{-1} **c** 0.71 s

d 4.0 m **e** 4.0 m s^{-1}

17 3.8 N **18** 1.0 m s^{-1}

19 Air resistance produces a drag force on the golf ball. Drag forces oppose the motion of the golf ball. This slows down the horizontal motion of the ball which reduces the range of the shot. Drag also reduces the vertical speed of the ball so that it stays it stays in the air for a shorter period of time and therefore travels a shorter distance, also reducing the range of the shot.

20 Air pressure: Generally, as a spinning ball travels in the air, it drags air around itself in the direction of the spin. The air moving past the ball also increases on one side and slows down on the other. Since faster moving air exerts less pressure, the air on the opposite direction of the ball is at a relatively greater pressure. It is this specific difference in pressure that creates the force which makes the ball to curve in flight.

Direction of ball movement: Whenever a ball is spinning through the air, the Magnus 'force' will push it in a direction perpendicular to the direction of movement.

Top spin:

- Magnus effect causes slower moving air to create high pressure above the ball and faster moving air creates low pressure underneath the ball.
- The direction of ball travel is always from high to low pressure.
- So the Magnus Force causes the ball to dip as it travels and the distance travelled is decreased from the non-spinning flight path.

Unit 2 Area of Study 2 Review

1 B **2** C **3** A **4** D **5** D

6 **a** $\frac{2}{3}$ pc **b** 137 510 AU

7 **a** The event horizon is a distance (a radius) around a black hole from which it is impossible for anything, even light, to escape. i.e. the escape velocity is faster than the speed of light

b 22.8 km to 59.3 km

8 Einstein's combined the three dimensions of space and movement in time to form the idea of spacetime. These four dimensions are all interdependent. Movement through space is not independent of moving through time; e.g. the faster you move through space, the slower you move through time. Core to Einstein's spacetime theory is that the speed of light will be the same for all observers.

9 The absolute magnitude of a star is the magnitude (brightness) that would be observed if the star were a distance of 10 pc from the Sun. The apparent brightness of a star is the brightness of a star as observed from Earth. Apparent brightness utilises a scale (devised by Hipparchus) ranging from –25 (very bright, e.g. our Sun) to +25 (very faint).

10 **a** 3.91×10^{26} W

b Luminosity would be four times greater.

c Luminosity would be sixteen times smaller.

11 D **12** D **13** C **14** D **15** A

16 B **17** B **18** C **19** A **20** C

21 Compressive forces: Forces applied in opposing directions cause compression of the material, forcing atoms closer together than they would be under equilibrium conditions.

Tensile forces: Forces exerted in opposite directions stretch a material, moving the atoms in the material further apart than they would be under equilibrium.

Shear forces: These occur when two or more forces act, but do not act in a line with each other. Shear forces tend to change the shape of an object without necessarily altering its dimensions.

22 332 kPa **23** 0.89 m **24** 40 Nm **25** 320 N

26 Carrying a long pole can lower the centre of mass of the tight rope walker, effectively improving their stability.

27 0.04 mm **28** 20 GPa **29** 10.8 J

30 A **31** B **32** C **33** D **34** C

35 ${}^{2}_{1}H + {}^{2}_{1}H \rightarrow {}^{3}_{2}He + {}^{1}_{0}n$ + energy

36 5.33568×10^{-13} J **37** 8.004×10^{13} J

38 **a** ${}^{238}_{92}U + {}^{1}_{0}n \rightarrow {}^{239}_{92}U$

b ${}^{239}_{92}U \rightarrow {}^{239}_{93}Np + {}^{0}_{+1}\beta$ ${}^{239}_{93}Np \rightarrow {}^{239}_{94}Pu + {}^{0}_{-1}\beta$

39 fast breeder reactor

40 C **41** B **42** B **43** D **44** B

45 **a** low-energy electromagnetic waves

b radiotracers emitting γ rays

c γ rays

d X-rays

46 40 mSv

47 **a** ${}^{124}_{54}Xe + {}^{1}_{0}\beta \rightarrow {}^{123}_{55}Cs + 2{}^{1}_{0}n$

b ${}^{123}_{55}Cs \rightarrow {}^{123}_{54}Xe + {}^{0}_{+1}\beta$

c ${}^{124}_{54}Xe \rightarrow {}^{123}_{53}I + {}^{0}_{+1}\beta$

48 **a** 1.25 kBq

b Not only does the sample decay, but the body will also rid itself of some of the I^{-123} through its normal mechanisms for removing waste.

c 250 μGy

49 Somatic effects of radiation exposure include the damage to body cells (any cells other than the sex cells). Symptoms of somatic effects can vary greatly, depending on the size of the dose. Symptoms include nausea, vomiting, diarrhoea, skin rashes, hair loss, drop in white blood cell levels.

Genetic effects result from exposure and consequent damage to the sex cells (sperm and ova). If the sex cells are damaged (changes made in the DNA of these cells from ionising radiation) then there maybe mutations in future generations.

50 D **51** A **52** B **53** B **54** C

55 Linac produces and accelerates electrons in a straight line.

Booster ring bends the electrons into a circular path using magnets and accelerates the electrons further. The energy is boosted through the use of an RF chamber.

Storage ring is where the electrons orbit. The electrons are travelling at speeds close to the speed of light. As they are 'bent' by the large magnetic fields, they emit their synchrotron light. Electrons can orbit for hours in the storage ring.

Beamline is the path that the synchrotron light takes from where it is produced (in the storage ring) to where it used to undertake experimental work.

56 The ATLAS experiment at the Large Hadron Collider was responsible for detection of the Higgs boson. The Higgs boson is a subatomic particle that essentially gives mass to all matter. The Standard Model in physics acknowledges the presence of four forces: gravitational force, electromagnetic force, strong nuclear force and weak nuclear force. The Standard Model has been able to link three of the four forces, but not the gravitational force. Why particles had mass could not be predicted from the Standard Model. The presence of a Higgs field and a subatomic particle, the Higgs boson, was theorised to explain how particles obtain mass. The detection of the Higgs boson was confirmation of this theory.

57 Synchrotron light is the light (or electromagnetic radiation) produced when charged particles are accelerated. As electrons are 'bent' and forced to travel in a circular path in the storage ring, they are continually being accelerated (their velocity is changing) and therefore they emit synchrotron light.

Synchrotron light has a number of key characteristics including:

- very high intensity
- a very broad spectral range (from infrared to X-rays)
- it is very collimated and experiences very low divergence
- it is emitted in very short pulses (less than a nanosecond)
- it is highly polarised.

58 Particle accelerators are being used in applications such as:

- cancer treatment
- carbon dating
- art restoration
- pharmaceutical research.

59 **a** 3.2×10^{10} m s^{-1}

b This value is greater than the speed of light, and nothing can travel at speeds greater than this.

c As particles are accelerated, their effective mass increases (Einstein's mass–energy equivalency). As particles are accelerated to speeds close to the speed of light, their effective mass increases significantly. The energy supplied to the electron by the synchrotron increases its speed and also its effective mass. The speed of the electron does not exceed the speed of light.

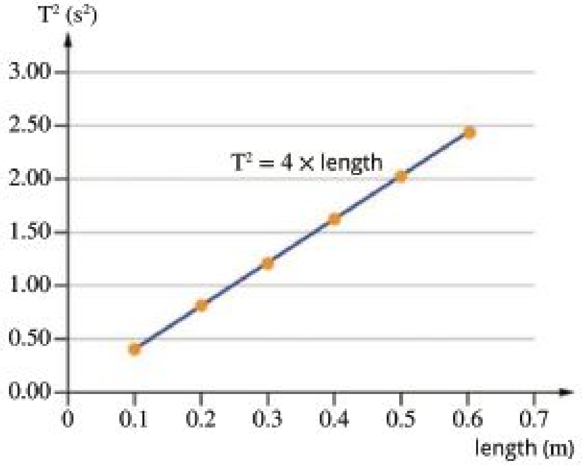
60 B **61** D **62** C **63** C **64** D
65 D **66** A **67** B **68** 0.71 **69** 0.5 m s^{-1} east.
The collision is considered elastic because the initial kinetic energies of both balls is equal to the final kinetic energy of both balls.

70 An inelastic collision is one in which kinetic energy is not conserved; that is, the kinetic energy before the collision is greater than the kinetic energy after the collision.
Nearly all collisions involve the conversion of kinetic energy to other forms such as heat and sound. As energy is always conserved in any interaction within a closed system, the final kinetic energy must therefore be less than the initial kinetic energy, and hence the collision an inelastic collision.

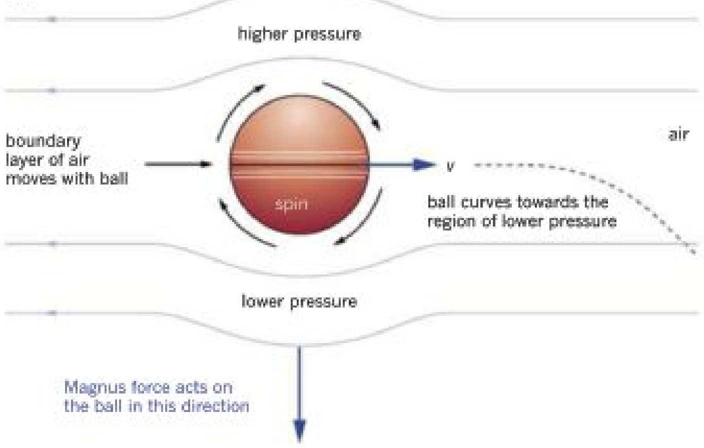
71 0.50 m
72 2.55Hz
73 a

Length of pendulum (m)	Time for 10 swings (s)	Period (s)	Period^2 (s^2)
0.1	6.3	0.63	0.40
0.2	9.0	0.90	0.81
0.3	11.0	1.10	1.21
0.4	12.7	1.27	1.61
0.5	14.2	1.42	2.02

b $g = 9.8$ m s^{-2}



74



75 200 g ball: terminal velocity is 18.8 m s^{-1}
400 g ball: terminal velocity is 26.6 m s^{-1}
The terminal velocity of the 400 g ball is $\sqrt{2}$ times greater than the terminal velocity of the 200 g ball.

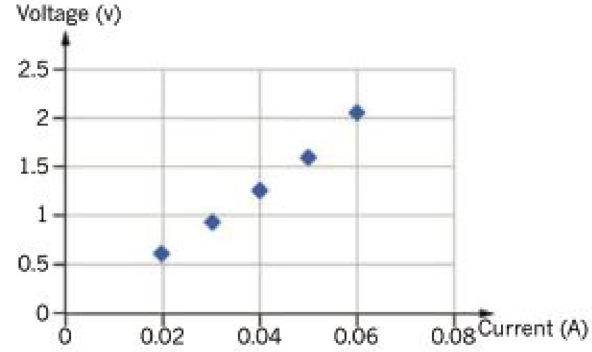
Chapter 19 Practical investigation

19.1 Designing and planning the investigation

1 a If the voltage is measured in units of number of batteries then it is a discrete value.
b If the voltage is measured with a voltmeter then the voltage would be continuous.
2 qualitative
3 C. A hypothesis should test only one independent variable and it should predict the relationship between the independent and dependent variable.
4 a valid **b** reliable **c** accurate
5 the type of grip

19.2 Conducting investigations and recording and presenting data

1 a systematic error **b** random error
2 two significant figures
3 a 22.7 **b** 19 **b** 22.5
4



5 Add a trend line or line of best fit.

19.3 Discussing investigations and drawing evidence-based conclusions

1 the proportional relationship between two variables
2 an inversely proportional relationship
3 directly proportional
4 time restraints and limited resources
5 An increase in current from 0.03 to 0.05 A produced an increase of 0.88 V across the resistor.

Chapter 19 Review

1 A prediction, based on evidence and prior knowledge, to answer the research question. It often takes the form of a proposed relationship between two or more variables.
2 dependent variable: flight displacement
independent variable: release angle
controlled variable: (any of) release velocity, release height, landing height, air resistance (including wind)
3 a the acceleration of the object
b the vertical acceleration of the falling object
c the rate of rotation of the springboard diver
4 elimination, substitution, isolation, engineering controls, administrative controls, personal protective equipment.
5 6.8 ± 0.4 cm s^{-1} **6** the mean
7 an exponential relationship
8 This graph should show a straight line with a positive gradient.
9 Any issues that could have affected the validity, accuracy, precision or reliability of the data plus any sources of error or uncertainty.
10 Bias is a form of systematic error resulting from a researcher's personal preferences or motivations.

Glossary

A

absolute brightness The actual brightness of a star. The absolute brightness of a star or other celestial object corresponds to the brightness measured in magnitudes of the object if it was 10 parsecs (32.6 light years) from Earth.

absolute magnitude The absolute magnitude of a star or other celestial object corresponds to the apparent magnitude of the object if it was 10 parsecs (32.6 light years) from Earth.

absolute zero The coldest possible temperature.

absorbed dose The amount of ionising radiation absorbed per kilogram of irradiated material, measured in grays (Gy).

absorption The taking up and storing of energy, such as radiation, light or sound, without it being reflected or transmitted. During absorption the energy may change from one form into another. When radiation strikes the electrons in an atom, the electrons move to a higher orbit or state of excitement by absorption of the radiation's energy.

acceleration The rate of change of velocity. Acceleration is a vector quantity. The SI unit for acceleration is m s^{-2}.

accretion The process where a black hole draws matter towards it, for example from a star or from a cloud of dust and gas.

actinide Any element of the series of 15 metallic elements from actinium (atomic number 89) through to lawrencium (atomic number 103) in the periodic table. They are all radioactive. The heavier members are extremely unstable and not of natural occurrence. They are a waste product of nuclear fission reactors.

activity The number of nuclei of a radioactive substance that decay each second, measured in becquerels (Bq).

air resistance The retarding force (drag) caused by collisions between air and moving objects.

alpha particle A particle consisting of two protons and two neutrons ejected from the nucleus of a radioactive nuclide.

alternating current In an alternating current (AC), electrons oscillate backwards and forwards around a mean position, as opposed to direct current (DC). Household power supplies usually operate at 240 V AC.

ammeter An ammeter is an instrument used to measure the electric current in a circuit. Electric current is measured in amperes (A), which is why it is called an ammeter.

angular speed A measure of how quickly an object is turning. Measured in radians per second (s^{-1}).

anisotropic Not uniform in all directions, showing different material properties in different directions.

annihilation The process in which matter is completely converted into energy. This is not a chemical process in which matter in one form is converted to matter in another form, as in burning.

anode A positively-charged electrode, as of an electrolytic cell, storage battery, or electron tube. Also, the negatively charged terminal of a primary cell or of a storage battery that is supplying current.

antimatter Particles that have the same mass as their ordinary matter equivalents but opposite properties like electromagnetic charge, spin, baryon number and lepton number.

antineutrino A neutral subatomic particle that interacts very weakly with other matter; the antimatter particle of neutrino.

apparent brightness How bright a star or other celestial object appears from Earth measured in magnitudes. It is based on the brightest star in the northern sky being magnitude +1. A change of −1 corresponds to a brightening by about 2.5 times.

apparent magnitude An arbitrary scale based on the brightest star in the northern sky being magnitude +1. A change of −1 corresponds to a brightening by about 2.5 times.

artificial transmutation The changing of one element or isotope into another. This happens during radioactive decay and during neutron bombardment in a nuclear reactor.

atomic number The number of protons in a nucleus.

axis of rotation An imaginary line through the centre of mass or pivot point of an object, which is perpendicular to the plane of rotation of the object.

B

baryon A composite particle composed of three quarks. Baryons belong to the particles called hadrons. The most common examples are the proton and the neutron.

baryon number A quantum number conserved in particle interactions. This means the sum of the baryon numbers before an interaction is equal to the sum of the baryon numbers after the interaction. Baryons (particles containing three quarks) are assigned +1, antibaryons are assigned −1 and all other particles are assigned 0.

base The support for a structure. For example, the base of a car is a rectangle with each of the four tyres at the corners of the rectangle.

beamline A pathway in which the photon beam generated in a synchrotron travels from the storage ring to an experiment room or end station.

beta particle An electron or positron ejected from the nucleus of a radioactive nuclide.

big bang theory The leading model for how the universe was created. It describes the universe starting in a high-density state and then expanding.

binary star A pair of stars that orbit a common centre of mass.

binding energy Energy required to split a nucleus into its separate nucleons.

black dwarf A black dwarf star is a white dwarf star that has cooled such that it no longer emits any significant heat or light.

black hole A collapsed star so massive that not even light can escape from its gravitational field.

booster ring Part of a synchrotron where the electron energy and speed is further increased.

brittle Describes a material that only displays elastic behaviour and fails at its yield point.

C

cantilever A beam that extends out horizontally beyond its supporting structure.

carbon footprint A measure of the human impact on the environment. The quantity of greenhouse gases released due to human activity.

cathode ray tube Vacuum tube where an electron beam is focused onto a screen.

centre of gravity The position from which the entire weight of the body or system is considered to act; the position at which the body will balance.

centre of mass A single point in an object where the mass can be considered to be 'concentrated' for the purposes of analysing motion.

chain reaction A series of nuclear fissions that may be controlled or uncontrolled.

charge A property of matter that causes electric effects. Protons have positive charge, electrons have negative charge and neutrons have no charge.

chemotherapy The use of drugs to destroy or slow the rapid growth of cancer cells in the body. These drugs are highly toxic to the cells and therefore may cause secondary side effects due to damage of surrounding healthy cells. Chemotherapy can use one drug or a combination of drugs. It can also use drugs that emit radiation (radiopharmaceuticals).

circuit breaker A device that automatically switches off an excessive current by detecting the magnetic field associated with it.

coefficient of restitution (COR) The ratio of the speed of a ball directly after a bounce to its speed before a bounce. COR is given by the rule: where v_1 is the speed before the bounce and v_2 is the speed after the bounce.

collinear Lying on the same straight line.

colour charge A property of quarks that is related to how they bond together. (Note, this is not related to the normal interpretation of colour.)

components The components of a force are two vectors at right angles to each other that when added together will be equivalent to the original force.

composite Material composed of more than one substance, employing favourable properties of each; for example, steel reinforced concrete.

compressive force Force acting in a material that has been squeezed or squashed.

conduction The movement of energy (such as heat) from one object to another without the net movement of particles (atoms or molecules).

conductor A substance, body or system that readily conducts heat, electricity, sound or light.

conservation law Describes a condition that something must remain unchanged. An example in particle physics is the conservation of charge. The sum of the charge on the particles before an interaction must be equal to the sum of the charge on the particles after the interaction.

conserved When a quantity that exists before an interaction is exactly equal to the quantity that exists after the interaction.

contact forces Forces that exist when one object or material is touching another. Friction, drag and normal reaction forces are contact forces.

control rod Material, commonly boron, steel or cadmium, that absorbs neutrons in a nuclear reactor.

controlled variable A variable that must be kept constant during an investigation.

convection A process of heat transfer through a gas or liquid by bulk motion of hotter material into a cooler region.

conventional current Basically the same as electric current. Conventional current is in the opposite direction to electron flow.

coolant A substance, commonly water, carbon dioxide or liquid sodium, used to transfer thermal energy from the core of a nuclear reactor.

core Part of a nuclear reactor where nuclear fission occurs and thermal energy is produced.

coulomb The SI unit of charge; 1C is equivalent to the combined charge of 6.2×10^{18} protons.

critical mass The minimum amount of enriched fissile material in the shape of a sphere that leads to a sustained fission reaction.

current The net flow of electric charge. Current is measured in amperes (A) where $1\text{A} = 1\text{C s}^{-1}$. By convention, electric current is assumed to flow from positive to negative.

cyclotron A particle accelerator device that accelerates particles outward from the centre of their trajectory along a spiral path.

D

daughter nucleus A nucleus on the product side of nuclear equation that results when a nucleus undergoes fission or radioactive decay.

decay series A sequence of radioactive decays that results in the formation of a stable isotope.

dependent variable The variable that may change in response to a change in the independent variable. On a graph, the dependent variable is plotted on the vertical axis.

diagnostic imaging A group of techniques and medical procedures needed to create visual representations of the organs and tissues that make up the human body in order to assess their health status. It includes techniques such as X-ray radiography, CT, PET, SPECT and MRI scans.

diffract The process affecting light and other wave forms that causes the wave to spread out as the wave passes through a narrow aperture or past an edge.

diffraction grating A polished glass surface with a large number of very fine etched grooves or slits designed to diffract light in order to create an optical spectrum.

dimension Space can be considered to consist of three length dimensions. These length dimensions are arranged at 90 degrees to each other with their point of intersection being the origin. The position of an object can be defined in relation to its position along each of the three dimensions. Typically, these three dimensions are labelled x, y and z. However, up-down, left-right and backward-forward are also appropriate.

dimensional analysis Using the units in a graph or formula to check that the derived term is correct.

diode A semiconductor device that has the special property that it will only allow electrical current to flow in one direction through it.

direct current In a direct current (DC), electrons travel in one direction only, as opposed to alternating current (AC). Batteries and electric cells provide direct current.

direction conventions Standardised systems for describing the direction in which an object is travelling. The use of cardinal points of a compass (N, S, E and W) is an example of a direction convention.

displacement An object's change in position, relative to its starting position and final position. Displacement does not consider the route the object took to change position, only where it started and where it ended. Displacement is a vector quantity. It is measured in metres (m) and given the symbol s.

distance travelled How far an object travels during a particular motion or journey. Distance is a scalar value. Direction is not required when expressing magnitude. It is measured in metres (m) and given the symbol d.

DNA Deoxyribonucleic acid is a type of nucleic acid that stores all the genetic information of an organism. The molecule exists as a double helix inside the cells of all living organisms.

Doppler effect A change in the observed frequency of a wave, such as sound or light that occurs when the source and observer are in motion relative to each other.

dose equivalent A measure of the biological damage inflicted on a tissue due to absorption of a defined quantity of radiation. Dose equivalent measurements take into account the nature of the radiation applied. It is measured in sieverts (Sv).

drag A retarding force experienced by any object that moves through a fluid like air or water.

ductile Describes a material that can withstand stresses greater than the elastic limit and can undergo significant plastic deformation before failing. Often used to describe metals.

E

earth The third wire (usually green or green and yellow) in electrical devices that acts as an important safety feature by carrying excess current due to a device malfunction directly into the Earth.

effective dose Measurement of the amount of radiation a tissue or organ has been exposed to which takes into account the sensitivity of the tissue towards the particular type of radiation. It is calculated by multiplying the value of dose equivalent by the 'weighting factor' (W). If more than one organ has been exposed to a radiation source then the total value of the effective dose is the sum of all the effective doses for all of the organs/tissues involved.

effective resistance A single resistance that could be used to replace a number of individual resistors for the purpose of circuit analysis.

effort Force applied to a lever to overcome the load.

elastic Material that returns to its original shape after being deformed.

elastic collision Collision in which kinetic energy is conserved.

elastic limit The yield point of a material, after which it either shows plastic behaviour or fails.

elastic potential energy Stored energy in a stretched or compressed material, measured in joules (J).

elastic region The linear section of a stress–strain graph. This region ends at the elastic limit, and so long as a material remains within the elastic region it will return to its original dimensions when the applied stress is removed.

electric circuit A continuous conducting loop that allows electric current to flow.

electric shock Also known as electrocution, in which excess electricity flows into the human body due to a device malfunction or electrical accident.

electrical potential energy Potential energy due to the concentration of charge in part of an electric circuit.

electricity A form of energy resulting from the existence of charged particles (electrons or protons). Electricity is fuelled by the attraction of particles with opposite charges and the repulsion of particles with the same charge.

electromagnetic radiation A wide range of frequencies (or wavelengths) that can be created by accelerating charges, which result in a rapidly changing magnetic field and electric field travelling out from the source.

electromagnetic spectrum The entire range of electromagnetic radiation. Consists of radio waves, microwaves, infrared radiation, visible light, ultraviolet, X-rays and gamma rays. In a vacuum, all electromagnetic radiation travels at $3.0 \times 10^{8}\ \text{m s}^{-1}$.

electron A negatively charged particle in the outer region of an atom; it can move from one object to another, creating an electrostatic charge. When electrons move in a conductor, they constitute an electric current.

electron flow The net flow of electrons. Although electric current is assumed to flow from positive to negative, electrons physically move from negative to positive.

electronvolt (eV) A small unit of energy. One electronvolt (1 eV) is the energy an electron would gain when accelerated across a potential difference of one volt: $1\ \text{eV} = 1.6 \times 10^{-19}\ \text{J}$.

electroweak theory The mathematical theory that describes both the electromagnetic and weak forces of nature. This theory is a component of the Standard Model of particle physics.

elementary charge The magnitude of the charge on an electron or proton: $e = 1.6 \times 10^{-19}\ \text{C}$.

emission spectrum The particular spectrum of frequencies light, unique to an element, emitted by an excited atom as the excited electrons drop back to lower energy levels.

emissivity The effectiveness of the surface of a material at emitting energy as thermal radiation. Emissivity is between 0 and 1 and takes into account the type and colour of the surface. Matt black surfaces have a value of e close to 1. A perfect emitter is termed a 'black body' with an e value of 1. Shiny surfaces have an e value close to 0.

emit Give out. Energy can be emitted in the form of heat, light, radio waves etc.

energy An object possesses energy if it has the ability to do work. Energy takes many forms, for example kinetic energy and potential energy.

energy levels The orbital levels in which electrons orbiting the nucleus of an atom can remain stable.

enhanced greenhouse effect Also known as climate change or global warming, is the impact on the climate from the additional heat retained due to the increased amounts of carbon dioxide and other greenhouse gases in the lower atmosphere.

evaporation The changing of a liquid into a gas, often under the influence of heat (as in the boiling of water).

event horizon The region, or limit, around a black hole marking the boundary inside which the gravitational force is strong enough to prevent all matter and radiation from escaping.

F

fast reactor A type of reactor in which the neutrons causing fission are not slowed by any moderator. The use of fast neutrons allows the production of more fissile nuclei than are destroyed, transforming U-238 into a fissile P-239 nucleus. This process is known as breeding, as it produces fissile P-239 nuclear fuel.

fermion A fermion can be an elementary particle, such as those that make up atoms, quarks and electrons, or it can be a composite particle, such as protons and neutrons. Bosons are particles that are not fermions. According to the spin-statistics theorem, particles with integer spin are bosons, while particles with half-integer spin are fermions.

fissile Capable of undergoing nuclear fission after capturing low-energy neutrons.

fission When a nucleus splits into two or more pieces, usually after bombardment by neutrons.

fission fragments Nuclides formed during nuclear fission; these are usually radioactive.

force A vector quantity which measures the magnitude and direction of a push or a pull. It is measured in newtons (N).

force arm The perpendicular distance between the axis of rotation and the line of action of the force.

force multiplier A lever in which the distance between the effort force and the pivot point is greater than the distance between the load and the pivot point, which means the effort force required is less than the load.

fossil fuel A natural fuel such as coal or gas, formed in the geological past from the remains of living organisms.

free fall The motion of a falling body under the effect of gravity only. No air resistance or propulsive forces are acting.

frequency A measure of the rate at which something occurs, for example the number of vibrations or cycles that are completed per second or the number of complete waves that pass a given point per second. Measured in hertz (Hz).

fuel rod Long, thin rod of enriched uranium used in a nuclear reactor.

fuse A circuit device that melts when too much current flows through it, breaking the circuit in the process and protecting the other circuit components.

fusion A process taking place inside stars in which small nuclei are forced together to make larger nuclei. Energy is released in the process.

G

gamma ray High-energy electromagnetic radiation ejected from the nucleus of a radioactive nuclide.

gauge boson Gauge bosons are force-carrier particles which, according to the Standard Model of particle physics, mediate the four fundamental forces.

Geiger counter A device for measuring radioactive emissions.

genetic Refers to the characteristics or modifications introduced to the DNA in the cell which are passed onto the offspring by sexual reproduction.

geothermal energy Energy produced by the internal heat of Earth.

giant A very large, bright non-main sequence star. Super giants are the very largest stars, being thousands of times brighter than the Sun but with a much shorter life time due to the faster rate of fusion.

gluon Gluons are elementary particles that act as exchange particles for strong nuclear force between quarks, similar to the exchange of photons in the electromagnetic force between two charged particles. Gluons themselves carry the colour charge of the strong interaction. Gluons can be considered to be the fundamental exchange particle underlying the strong interaction between protons and neutrons in a nucleus.

gravitational force The force of attraction of all matter on all other matter due to mass.

gravitational potential Energy available to an object due to its position in a gravitational field. Measured in joules (J).

greenhouse effect The trapping of the Sun's energy in a planet's atmosphere, which warms the planet. Thermal radiation from a planetary surface is absorbed by atmospheric greenhouse gases, and is re-radiated in all directions.

greenhouse gas A gas that contributes to the greenhouse effect by absorbing infrared radiation. Carbon dioxide and chlorofluorocarbons found in aerosols are examples of greenhouse gases.

H

hadron A composite particle that contains quarks held together by the strong force. Hadrons are subdivided into two families: baryons (e.g. the proton and neutron) and mesons (the pion and kaon).

half-life The time taken for half of the nuclei of a radioactive isotope to decay.

hard X-rays High-energy X-rays with a wavelength less than 0.1 nm and energy values greater than 10 keV. They have high penetrating capacity and frequency. They are commonly used for radiation therapy purposes.

heat The energy transferred from a hotter object to a cooler one that increases the kinetic and/or potential energy of the particles in the cooler object.

heat exchanger Part of a nuclear reactor where heat drawn from the reactor core is used to turn water into steam.

heavy water Water that has a higher than normal proportion of water molecules that contain deuterium.

Hertzsprung–Russell (H–R) diagram A plot of the luminosity of stars against surface temperature that classifies stars by types.

Higgs boson Elementary particle discovered in 2012 at CERN that essentially gives mass to all elementary particles.

Hooke's law Elastic materials for which there is a direct relationship between the force acting on them and the extension or compression that they undergo are said to obey Hooke's law.

Hubble Space Telescope A telescope launched by NASA in 1990. It has a 2.4-m mirror and can 'see' in the ultraviolet and infrared as well as visible wavelengths.

Hubble Constant The unit of measurement used in Hubble's law, which describes the expansion of the universe. It has a value of around $70\ \text{km s}^{-1}\ \text{Mpc}^{-1}$.

Hubble's law A law created by Edwin Hubble that states that the rate at which astronomical objects in the universe move apart from each other is proportional to their distance from each other.

hydronic heating A cooling or heating system that uses circulating water to transport heat.

hydrostatic equilibrium For the majority of the life of a star, the gravitational force from the mass of the star is in balance with the gas pressure due to energy generation in the core of the star. The star is 'hydrostatic equilibrium' during this phase of its life.

hypothesis A proposed explanation for an observed phenomenon.

hysteresis A memory effect where the properties of the material depend on its prior treatment.

I

impulse The change in momentum of an object is also called the impulse of an object. The impulse is calculated by the final momentum minus the initial momentum.

incident Arriving at or striking a surface, especially a beam of light or radiation, or particles.

independent variable The variable that is selected and deliberately changed by the researcher. On a graph, the independent variable is plotted on the horizontal axis.

inelastic collision Collision in which kinetic energy is not conserved.

inertia A property of an object, related to its mass, that opposes changes in motion.

insulator A material or an object that does not easily allow heat, electricity, light or sound to pass through it. Air, cloth and rubber are good electrical insulators; feathers and wool are good thermal insulators.

interference The variation of wave amplitude that forms when two or more waves of the same or different frequencies come together. The amplitude of the resulting wave will be either larger or smaller than the amplitude of the individual waves, depending on whether or not their peaks and troughs match up. If the peaks of the waves match up, the amplitude of the resulting wave will be larger than that of the individual waves. This is called constructive interference. If the peaks and troughs of the individual waves do not match up, the resulting amplitude is smaller. This interference is called destructive interference.

interference pattern The superposition of two or more electromagnetic waveforms to form a resultant wave in which the displacement varies as the component waves either reinforce or cancel the wave movement.

interglacial A period of milder climate between two glacial periods.

internal energy The total kinetic and potential energy of the particles within a substance.

intrinsic brightness The actual brightness of a star, regardless of the distance from the observer.

ion Atom of a chemical element in which the number of electrons and protons is not equal and therefore the atom is electrically charged. If extra electrons are present, the ion has a negative charge. If electrons are missing, the ion has a positive charge.

ionising radiation Radiation with enough energy to alter the molecular structure of matter by displacing one or more electrons from an atom and thus creating electrically charged ions.

isotope Atoms with the same number of protons but with different numbers of neutrons.

isthmus A narrow strip of land with sea on either side, forming a link between two larger areas of land.

J

junction A point in an electric circuit from which current can flow into or out of from more than one direction.

K

kelvin An absolute temperature scale based on the triple point of water.

kilowatt hour (kWh) Unit of energy equivalent to 3.6 megajoules. The equivalent amount of energy as a 1000 W device turned on for one hour. It is the unit of measure of electricity usage that is measured by electricity meters and appears on electricity bills.

kinetic energy The energy of a moving body, measured in joules (J).

kinetic particle model A model that states that the small particles (atoms or molecules) that make up all matter have kinetic energy, which means that all particles are in constant motion, even in solids.

L

laminate Material that is composed of two or more materials bonded together in a layered structure.

Large Hadron Collider The world's largest and most powerful particle accelerator, located at CERN laboratories near Geneva, in Switzerland. It is a 27 km highly evacuated tube in which particles are accelerated to 99.999999% of the speed of light. They are held in place in the ring by a huge array of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way.

latent heat The 'hidden' energy used to change the state of a substance at the same temperature, i.e. the energy is not seen as a change in temperature.

latent heat of fusion The energy required to change 1 kg of solid to a liquid at its melting point.

latent heat of vaporisation The energy required to change 1 kg of liquid to a gas at its boiling point.

lepton Leptons are the six fundamental particles of the Standard Model that do not experience the strong force. Examples are the electron, the muon and neutrino. The charged leptons experience the electromagnetic force but the neutral neutrinos do not.

lever A simple machine consisting of a beam that rotates about a pivot point when a torque is applied.

light dependent resistor (LDR) A non-ohmic device whose resistance varies with the light intensity that falls upon it.

light emitting diode (LED) A type of diode that emits light as current passes through it.

linac Abbreviation for 'linear accelerator'. These are subatomic particle accelerator devices that are used to generate beams of high-energy X-rays to be used in radiation therapy for cancer treatment.

line of action of the force The line along which a force is acting. The line of action extends forwards and backwards from the force vector.

linear accelerator Type a particle accelerator in which particles are accelerated in a straight line.

lithography The process of printing a pattern onto another surface.

load A force, often a weight, on a lever or material.

luminosity The absolute brightness of a star measured in watts of total energy output.

M

magnitude The size or extent of something, with no need for direction. In physics, this is usually a quantitative measure expressed as a number of a standard unit.

main sequence A group of stars lying on a line running from the top left to the bottom right of the Hertzsprung–Russell diagram.

mantle The interior section of Earth, found between the crust and the outer core.

mass An amount of matter. One kilogram of mass is equal to the amount comprising the standard kilogram cylinder of platinum–iridium. Mass can be defined by the amount of matter that would result in an acceleration of 1 m s^{-2} when a force of 1 N is applied in a frictionless environment.

mass number The number of nucleons (protons and neutrons) in a nucleus.

mean The average value that is calculated by taking the sum of all values and then dividing by the total number of values.

mechanical advantage Ratio of the load force to the effort force for a lever. If the mechanical advantage is greater than one, then the lever is a force multiplier.

mechanical energy The energy that a body possesses due to its position or motion. Kinetic energy, gravitational energy and elastic potential energy are all forms of mechanical energy.

median The middle piece of data when a data set is listed in order.

meson Mesons are unstable subatomic particles composed of one quark and one antiquark. They are in the hadron family–particles made of quarks. Baryons–subatomic particles composed of three quarks–are also part of the hadron family.

metal Material in which some of the electrons are only loosely attracted to their atomic nuclei.

mode The most common piece of data in a data set.

moderator A material, usually graphite or water, that slows neutrons in a nuclear reactor.

momentum The product of an object's mass and velocity. Objects with larger momentum require a larger force to stop them in the same time that an object with smaller momentum takes to stop. It is given by the equation: $p = mv$.

mutation Any change in the structure or composition of DNA, which in turn, alters the genetic information stored in a cell.

N

nebula An interstellar cloud in outer space composed of dust and gasses. The Latin word for cloud is *nebula*.

net charge When the number of positive and negative charges in an object is not balanced.

net force The vector sum of all the individual forces acting on a body.

neutral No electric charge, or a situation in which positive and negative charges are balanced.

neutral equilibrium A situation in which an object will remain stationary no matter where it is, for example a ball on a horizontal table.

neutrino An almost massless, neutral particle released during some nuclear reactions.

neutron An uncharged subatomic particle.

xneutron bombardment A physical process by which a stable atomic nucleus is bombarded with high-speed ions or neutrons inside a nuclear reactor. As a result, the atomic structure of the original nucleus changes and it becomes a different element.

neutron star The remnant of a supernova, consisting entirely of neutrons.

Newton SI unit of force. One newton (1 N) is defined as the force required to make a mass of 1 kg accelerate at 1 kg m s^{-2}.

Newton's First Law States that an object will maintain a constant velocity unless an unbalanced, external force acts on it.

Newton's Second Law States that force is equal to the rate of change of momentum. This can be processed mathematically to: the acceleration of an object is directly proportional to the force on the object and inversely proportional to the mass of the object.

Newton's Third Law States that for every action (force), there is an equal and opposite reaction (force).

non-contact forces Forces that act at a distance and do not require the bodies to actually touch each other. Gravitational, magnetic and electric forces are non-contact forces.

non-ionising radiation Radiation that does not have enough energy to break the molecular bonds within molecules and to alter the number of electrons in an atom. Lower forms of energy in the electromagnetic spectrum such as radio waves, microwaves, visible light and UVA radiation are non-ionising.

non-metal Material in which all of the electrons are strongly attracted to their atomic nuclei.

non-ohmic Not behaving according to Ohm's law; resistance changes depending on the potential difference.

normal reaction force The force provided by a surface that is perpendicular to the surface with which an object is in contact. It is important to note that the two reciprocal forces (provided by the object and provided by the surface) are not the pair of forces described in Newton's Third Law. This is because the two forces are both acting on the same object and Newton's Third Law force pairs always act on different objects.

nuclear fission reactor A device built to control and harness the energy from nuclear fission reactions. Nuclear power stations are one type of nuclear reactor; other uses include medical, industrial, weapons and research purposes. Australia has one nuclear reactor (OPAL) at Lucas Heights in Sydney.

nuclear fusion The process in which two light nuclei, such as hydrogen, are combined to form a larger nucleus and to produce light and heat.

nuclear power station A nuclear reactor built for the purpose of producing electrical power. Nuclear fission is used to generate heat, which creates steam to drive turbines, in much the same way as a coal-burning power plant. As at 2016, Australia does not have any nuclear power stations generating electricity.

nuclear transmutation The changing of one element into another.

nucleon A particle located in the nucleus of an atom.

nucleus The central part of an atom.

nuclide The range of atomic nuclei associated with a particular atom, which is defined by its atomic number, and the various isotopes of that atom as identified by the mass number.

O

outlier A value that lies outside the main group of data of which it is a part. Outliers in data could be caused by errors in the experiment.

overload When an unsafe amount of current flows through a wire; for example, when too many electrical appliances are connected to the same power point.

P

pair production The creation of a particle and its antiparticle. This is commonly the result of two photons interacting or a photon interacting with an atomic nucleus.

Pangaea A supercontinent that included all the world's landmasses in the late Paleozoic era. It formed approximately 300 million years ago and then began to break apart after about 100 million years. According to plate tectonics, it eventually broke up into the continents we know today.

parallax movement The apparent movement of the closer stars relative to the background stars which is actually due to the motion of Earth around the Sun.

parallel circuit A circuit that contains junctions; the current drawn from the battery, cell or electricity supply splits before it reaches the components and re-joins afterwards.

parent nucleus A nucleus on the reactant side of a nuclear equation that when struck by a neutron undergoes fission or simply decays by natural means.

parsec The distance to a star that has a parallax angle of one arcsecond; equal to 206 265 AU.

particle accelerator A device where charged particles are accelerated to near the speed of light in order to bombard a target under investigation.

particle radiation Refers to a form of energy that is given off by atoms, in the form of small, high-energy sub-atomic particles, such as alpha particles, beta particles and positrons.

particulate The tiny solid (or sometimes liquid) matter that is found in Earth's atmosphere, i.e. pollution. This is often released into the atmosphere as part of burning fossil fuels, such as carbon monoxide from car exhausts.

passive design Energy efficient designs that lead to little or no mechanical heating/cooling requirements in a building.

pendulum A mass hung by a string from a fixed support. When the mass is displaced from equilibrium, it swings back and forth experiencing simple harmonic motion.

penetrating ability A measure of how easily radiation passes through matter.

perforated solar collector wall A wall that is built on the winter sun side of a building. It uses the Sun's rays to warm air that circulates through a building. It provides a passive means of heating a building.

period The time interval for one vibration or cycle to be completed.

photo resist A photosensitive resin that loses its resistance to chemical etching when exposed to radiation. Used especially in the transference of a circuit pattern to a semiconductor chip during the production of an integrated circuit.

photon Small energy packets of electromagnetic radiation which have no mass, no charge and travel through a vacuum at the speed of light.

photosphere The luminous envelope surrounding the 'surface' of a star from which its light and heat radiate.

photovoltaic cell A cell that converts visible light from the Sun into direct current electricity.

pivot point A point about which an object can rotate.

planetary nebula The usually ring-shaped nebula formed by an expanding shell of gas around an ageing star. These are basically layers formed as a result of the various stages of fusion reactions and may involve a considerable proportion of the star's mass.

plastic region Part of a stress–strain curve where the material deforms significantly for little additional stress. Occurs after the material has passed its elastic limit.

point mass To simplify motion analysis, it can be assumed that the mass of an object is concentrated at a single point at its centre of mass.

position The location of an object with respect to a reference point. Position is a vector quantity.

positron The antimatter pair of the electron. This means it shares the same mass as an electron but has opposite properties like electromagnetic charge and spin.

potential difference The difference in electric potential between two points in a circuit; measured by a voltmeter when placed across a circuit. A battery creates the potential difference across a circuit, which drives the current.

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

potentiometer A circuit device consisting of a three-terminal resistor with a sliding or rotating contact called the wiper. It can also be connected at one end and at the wiper to create a variable resistor.

power The rate at which work is done; a scalar quantity measured in watts (W).

principle of moments In order for an object to be in rotational equilibrium, the sum of the moments in a clockwise direction must balance the sum of the moments in an anticlockwise direction.

prosthesis A man-made device designed to replace a body part.

proton A positively charged subatomic particle.

protostar A collapsing cloud of gas that will eventually become a star.

pulsar A celestial object that is thought to be the rapidly rotating neutron star left after the explosion of a giant star. It emits regular pulses of radio waves and other electromagnetic radiation as it revolves at rates of up to 1000 pulses per second.

Q

qualitative variable A variable that can be observed but not measured.

quality factor The number used to indicate the weighting of the biological impact of radiation.

quantitative variable A variable that can be measured.

quantum mechanics The physics of subatomic particles, including their behaviour and physical interactions.

quantum number The numbers that describe the properties of a particle, like charge and spin.

quark Quarks are six of the fundamental particles in the universe. They cannot be found individually and only exist bound by the strong force within hadrons.

R

radian The measure of a central angle subtending an arc equal in length to the radius; one full revolution.

radiation Rays or particles that carry energy. Also, the process by which energy is emitted by an object or system, transmitted through an intervening medium or space, and absorbed by another object or system.

radioactive Something that spontaneously emits radiation in the form of alpha particles, beta particles and gamma rays.

radioactive tracer A chemical compound or atom that emits radioactivity that can be used to trace the position and localisation of a molecule within an organ or tissue in the body.

radioisotope An isotope of a chemical element that emits radioactivity due to its unstable combination of neutrons and protons in the nucleus.

radiopharmaceutical One of a group of drugs that contains a radioactive tracer attached to it.

random error An error in measurement that occurs in an unpredictable manner.

raw data The actual measurements taken directly during an investigation without being processed in any way.

redshift The change (lowering) of frequency that occurs in any wave phenomenon as the source of the waves moves away from the observer. Likewise, a blueshift occurs if the source is moving towards the observer.

reference point A point about which a rotational equilibrium can be calculated for an object that is in static equilibrium. This point can be anywhere, but is best selected to cancel the torque of an unknown force in a problem.

refraction The bending of the direction of travel of a ray of light, sound or other wave as it enters a medium of differing density.

reliability The consistency of the results obtained from an experiment or collection of data. Reliable results are also repeatable, meaning another scientist performing the same analysis will come up with the same results.

residual current device (RCD) A device that can detect a difference in the active and neutral wires and switch off current in dangerous situations to help prevent electrocution.

resistance A measure of how much an object or material resists the flow of current; the ratio of the potential difference across a circuit component and the current flowing through it: $R = V/I$ Resistance is measured in ohms (Ω).

resistor A circuit component, often used to control the amount of current in a circuit by providing a constant resistance. Resistors are ohmic conductors, i.e. they obey Ohm's law.

resultant One vector that is the sum of two or more vectors.

rotational equilibrium A situation in which the sum of all the clockwise torques is equal to the sum of all the anticlockwise torques.

rotational kinetic energy The kinetic energy due to the rotation of an object.

rotational motion The motion of any object can be analysed in terms of translation (i.e. movement from one point in space to another) and rotation (i.e. turning around a central point).

S

scalar A physical quantity that is represented by magnitude and units only. Mass, time and speed are examples of scalar quantities.

Schwarzschild radius The distance to the 'event horizon' around a black hole, a distance from the singularity at which it is impossible for even light to escape.

scientific method The process scientists apply to construct theories to explain practical observations.

scientific theory The theory that results from a tested and supported hypothesis, as per the scientific method.

semiconductor A material that will conduct electricity only under particular conditions, placing its properties somewhere between those of a conductor and an insulator.

series circuit When circuit components are connected one after another in a continuous loop so that the same current passes through each component.

shear force If a sideways force acts across the top of a material, and an opposing sideways force acts across the bottom, then the material is under shear stress.

short circuit The situation in which a good conductor is inadvertently placed across a battery and an excessive current flows, which may cause damage.

significant figures The numbers in a measurement or calculation that convey meaning and precision.

simple harmonic motion Type of motion of an object, where its acceleration is directly proportional to its displacement from the equilibrium position. The acceleration of the object is also directed towards the equilibrium position.

soft X-rays X-rays with wavelengths greater than 0.1 nm. They have lower energy and penetrative power compared to hard X-rays. They are used mostly for diagnostic imaging purposes.

solar collector Transform solar radiation into heat energy and then transfer that heat to water, solar fluid or air. This energy can be used for hot water heating, space heating or even air conditioning.

somatic Refers to all cells in the human body, except those present in the reproductive organs (i.e. the ovaries and testes). Mutations affecting somatic cells are not passed onto offspring.

spacetime A term used to describe the situation in which the three-dimensional space coordinate system (x, y, z) is linked to the one-dimensional time system.

specific heat capacity The amount of energy that must be transferred to change the temperature of 1 kg of material by 1°C or 1K.

spectroscope A device, optical or digital, for producing and recording spectra from sources of electromagnetic radiation for examination.

spectroscopic parallax Misleading term describing the process of finding the distance to stars that are beyond the range of stellar parallax.

speed The ratio of distance travelled to time taken. Speed is a scalar quantity. The SI unit for speed is m s^{-1}.

speed multiplier A lever in which the distance between the effort force and the pivot point is less than the distance between the load and the pivot point. During rotation, the point of application of the load travels further than the point of application of the effort, and therefore the point of application of the load travels faster.

stable equilibrium A situation in which an object will return to its equilibrium position, even when it is displaced by a force. For example, a ball placed in a large bowl will always return to the bottom of the bowl.

standard candle A class of astronomical objects whose distances can be calculated by comparing their observed brightness with their known luminosity.

Standard Model The Standard Model of particle physics is a mathematical description of all known particles and three of the forces acting on them. It is currently the most successful theory for predicting the behaviour and properties of the particles that exist in nature.

static equilibrium A situation in which an object is in both translational equilibrium and rotational equilibrium.

steady state theory A model of the universe based on an idea called the perfect cosmological principle. This states that the universe on the largest scales looks essentially the same everywhere at all times. Therefore, the universe maintains the same average density of matter forever. The Steady State theory is not the accepted model of the universe.

stellar parallax The difference in direction of a celestial object as seen by an observer from two widely separated points. The two points generally used coincide with the position of Earth at opposite sides of the Sun during its annual rotation since that is the most widely separated locations available to astronomers.

storage ring Part of a synchrotron where the electrons are held for a long period of time while they continuously emit synchrotron light.

strain The change in length of a material relative to its unstressed length. Strain is dimensionless, and is generally reported as positive for compression or tension.

strain energy The work done in changing the length of a material.

strength The maximum stress that a material can withstand before failing under tension or compression. Strength is measured in newtons per square metre (N m^{-2}) or pascals (Pa).

stress The applied force per unit cross-sectional area in a material is called stress. Stress is measured in newtons per square metre (N m^{-2}) or pascals (Pa).

strong nuclear force A short-range but powerful force of attraction that acts between all the nucleons in the nucleus. The strong nuclear force acts on quarks and binds them together in hadrons. It also acts at larger distances to bind protons and neutrons together within atomic nuclei. In the Standard Model of particle physics the strong force is described by quantum chromodynamics and is mediated by an exchange of gluons.

subcritical mass A quantity of fissile material that is too small to sustain a chain reaction.

supercritical mass A quantity of fissile material that is large enough to sustain a chain reaction.

supernova A giant explosion that occurs when a star many times larger than our Sun runs out of nuclear fuel.

synchrotron A type of particle accelerator in which electrons are accelerated by fluctuating electric fields and bent by a series of magnets to follow a circular path. The accelerating electrons produce a spectrum of electromagnetic radiation known as synchrotron light.

synchrotron light The broad spectrum of electromagnetic radiation generated by a synchrotron ranging from infrared to X-rays.

systematic error An error that is consistent and will occur again if the investigation is repeated in the same way.

T

temperature A measure of the average kinetic energy of the particles in a substance. Temperature can be measured in degrees Celsius (°C) or kelvin (K).

terminal velocity Velocity at which the force of drag is equal to the weight of an object.

thermal contact When two objects are in contact so that energy exchange via heat transfer is possible.

thermal energy A form of energy transferred as a result of a difference in temperature or average kinetic energy within a system.

thermal equilibrium For two bodies in thermal contact, the point at which the two reach the same temperature and there is no further net transfer of thermal energy.

thermistor A non-ohmic device whose resistance varies with temperature.

torque A turning effect caused by a force acting along a line that is not directed through a pivot point or a centre of mass.

toughness The ability of a material to absorb energy before it fails. Tough materials undergo significant plastic deformation before failing.

transducer A device that receives a signal in the form of one type of energy and converts it into another form of energy.

transfer The conversion of energy from one system to another.

transform To change from one thing to another; for example, to change energy from electrical potential energy to kinetic energy.

translation The motion by which a body shifts from one point in space to another, as opposed to rotational motion, where the body rotates at a fixed point in space.

translational equilibrium A situation in which the sum of the forces acting on an object are equal to zero.

translational kinetic energy The energy due to motion from one location to another.

translational motion The motion of any object can be analysed in terms of translation (i.e. movement from one point in space to another) and rotation (i.e. turning around a central point).

Trombe wall A wall that is built on the winter sun side of a building with a glass external layer and a high heat capacity internal layer separated by a layer of air. A Trombe wall is used in passive solar building.

U

uncertainty The description of the range of data obtained; the maximum variance from the mean.

units Properties related to physical measurements. Units can be fundamental like metres (m), seconds (s) or kilograms (kg). Units can also be derived by combining fundamental units like metres per second (m s^{-1}).

unstable equilibrium A situation in which an object will accelerate and will not return to its equilibrium position when it is displaced by a force. For example, a sphere lying on top of a dome.

V

validity The reasonableness of the results received from an experiment or collection of data. Valid results meet all the requirements of the criteria of the scientific method.

variable A factor or condition that can change.

vector A physical quantity that requires magnitude, units and a direction in order to be fully defined. Velocity, acceleration and force are examples of vector quantities.

vector diagram A system of adding vectors where each vector is drawn head-to-tail, with the resultant vector drawn from the tail of the first vector to the head of the last vector.

velocity The ratio of displacement to time taken. Velocity is a vector quantity. The SI unit for velocity is $m\ s^{-1}$.

viscoelasticity Material exhibits both viscous and elastic properties. Viscous materials have a time dependent response to stress. For viscoelastic materials, the stiffness is dependent on the rate at which force is applied.

volatile Liquids with weak surface bonds that evaporate rapidly.

voltage divider A series circuit with two or more components and where the voltage supplied to the circuit is shared (or divided) between the components in the circuit.

volt The unit of electrical potential. One volt is equal to one joule of potential energy given to one coulomb of charge in a source of potential difference. The voltage (or the number of volts) is another name for the potential difference.

voltmeter A device used to measure the voltage change between two points in a circuit.

W

wavelength The distance between one peak or crest of a wave of light, heat or other energy and the next corresponding peak or crest (symbol: λ).

weak nuclear force The weak nuclear force is the interaction between quarks responsible for changing from one type of quark to another. This force is described by electroweak theory and is responsible for radioactive decay and nuclear fission. In the Standard Model of particle physics this force is mediated by the W and Z bosons.

weight force The force of attraction on a body due to gravity.

weighting factor The number used to indicate the different sensitivities towards radiation by each organ in the body (W).

white dwarf A small, very dense star around the size of a planet. A white dwarf is formed when a low-mass star, such as our Sun, exhausts all its fuel and loses its outer layers. It will eventually cool to become a black dwarf as no nuclear fuel remains to generate additional heat and energy.

work The transfer of energy as a result of the application of a force; measured by multiplying the force and the displacement of its point of application along the line of action. Measured in joules (J).

X

X-ray diffraction X-rays diffract from a crystalline structure to produce a particular pattern. Analysis of this pattern reveals information about the spacing of atoms in the sample.

Y

yield point The elastic limit of a material, after which it either shows plastic behaviour or fails.

Young's modulus The ratio of stress per unit strain for a material, giving an indication of stiffness or flexibility. The modulus is measured in pascals (Pa).

Index

Acknowledgements

We thank the following for their contributions to our text book. The following abbreviations are used in this list: t = top, b = bottom, l = left, r = right, c = centre.

123RF : Mikhail Abramov, p. 408cl; Pashkov Andrey, p. 385(a); arenaphotouk, p. 409; Mikalai Bachkou, pp. x, 1, 83-6; Mikolaj Cepil, pp. 257, 445-52; Dmitriy Denysov, pp. 553, 638-44; Anton Drozdov, pp. 645-7; Warren Faidley, p. 87; fineart, p. 597; Pawel Gaca, p. 385(b); grigvovan, p. 408tl; Dan Kosmayer, p. 604b; keng po leung, pp. 287, 330; lianem, p. 380b; Sean Nel, p. 341; nimon thong-uthai, p. 29; maigi, pp. 123, 161-6; Viktar Malyshchyts, p. 256; matteo NATALE, p. 141; mohd hairul fiza musa , p. ix; Dario Lo Presti, p. 331; serg_v, p. 168; Anna Vaczi, p. 531; verdateo, p. 376b; yuanyuan xie, p. 375.

AAP: Philippe Halsman, p. 229; Stephen Pond, p. 492.

Alamy Ltd: aberCPC, p. 517tr; Archive Pics, p. 33; Danny Callcut, p. 245cr; Ashley Cooper, p. 75; Glasshouse Images, p. 537l; Interphoto, pp. 133, 537r Photo Researchers, p. 232r; Science Picture Co, p. 507; Ron Sumners, p. 506; World History Archive, p. 474r; Yakoniva, p. 557b.

Alexander Turnbull Library: S P Andrew Ltd : Portrait negatives. Ref: 1/2-043316-F. Alexander Turnbull Library, Wellington, New Zealand, p. 226bc.

Amana Images: pp. 226br, 332bl, 458tl, 474l; 2/Ryan McVay/Ocean, p. 358r; adoc-photos, p. 458cl; Bettmann, p. 226c; Alfredo Escobar, p. 298; Mark Garlick, p. 481; Dr. Robert Gendler, p. 479t; Peter Ginter/Science Faction, p. 577; ITAR-TASS, p. 221b; Mainichi Newspaper/AFLO, p. 544l; Michael Nicholson, p. 514; Shocktrek Images, p. 480; Annette Soumillard/Hemis, p. 532c; Visuals Unlimited, pp. 332tl, 477.

Daniel G. Amen, MD: www.amenclinics.com, p. 572.

Australian Bureau of Meteorology: p. 62.

Australian Synchrotron: pp. 238, 581.

Tom Barbour: www.thermalimageuk.com, p. 5b.

Bielasko: p. 385(c).

Bruce Rankin Illustration: p. 626.

Keith Burrows: pp. 150, 154.

CERN: p. 592.

City of Melbourne: p. 45.

Climate Reanalyzer: Image obtained using Climate Reanalyzer (http://cci-reanalyzer.org), Climate Change Institute, University of Maine, USA, p. 57.

Commonwealth of Australia: Department of the Environment, Water, Heritage and the Arts, CC BY 4.0 Licence p. 46.

CSIRO: Reproduced by permission of CSIRO Australia, © CSIRO http://www.csiro.au/greenhouse-gases/, p. 60.

David Malin Images: Akira Fujii/David Malin Images, p. 462.

Michael Devereux: p. 621.

DK Images: Tim Ridley & Dorling Kindersley, p. 145.

Fairfax Photo Sales: Bryan O'Brien, p. 282r; Andew Quilty, p. 561.

FLIR Systems Inc: p. 26r.

Fotolia: Ilya akinshin, p. 417br; Am, p. 421t; Ammentorp, p. 394; Amphotolt, p. 102bl; Leonid Andronov, p. 245tr; Apops, p. 270t; Olesia Bilkei, p. 25; Yuri Bizgaimer, p. 412; Alison Bowden, p. 354; candy1812, p. 41; Corey Chestnut, p. 496; Constantinos, p. 240; Coprid, p. 146bl; Les Cunliffe, p. 362; Cutimage, p. 102tr; Sahua D, p. 427r; demarco, p. 90; destina, p. 421b; George Dolgikh, p. 107; Dreaming Andy, p. 151r; Ezume Images, p. 650; Michael Flippo, p. 5t; Bert Folsom, p. 111; Fotomatrix, p. 112tr; JM Fotografie, p. 105; Generalfmv, p. 200; Georgerudy, p. 607l; Deyan Georgiev, p. 112cr; gladophoto, p. 609; Francisco Javier Gil, p. 116; Warren Goldswain, p. 26l; M. Gove, p. 619br; ISTANBUL2009, p. 622br, ch.18 answers; Juulijs, p. 338t; Kadmy, p. 270b; KAR, p. 652; Key99, p. 605b; Oleg Koslov, p. 363; Dusan Kostic, p. 499; leungchopan, p. 95; manfredxy, p. 213; Marcel, p. 433tr; Elisheva Monasevich, p. 259; nobeastsofierce, p. 146tr; Viktar Ramanenka, p. 413l; Romano72, p. 334r; Boris Ryaposov, p. 631; Roman Samokhin, p. 112br; Ian Scott, p. 151l; Siegmar, p. 402; Silver, p. 18; Ljupco Smokovski, pp. 415, 416; somchaisom, p. 112l; Studioloco, p. 417tr; Stuporter, p. 421c; sumire8, p. 34; tadamichi, p. 662; Tagstock, p. 22; tharatheppti, p. 40br; T-REX, p. 93l; TTstudio, urcis, p. 199; p. 30; Kim Warden, p. 334l; William87, p. 417bl.

Getty Images: AFP/Stringer, pp. 282l, 563; American Institute of Physics, p. 579; The Asahi Shimbun, p. 567b; Alex Bartel, p. 533; Adrian Dennis/AFP, p. 525t; Michael Dodge/Stringer, p. 366; Julian Finney, p. 369; Gallo Images/Stringer, p. 634; Stuart Hannagan, p. 291b; L. Livermore Nat Lab, p. 525br; Bryn Lennon, p. 292; Library of Congress, p. 555; Dean Loomis, p. 303t; Novosti, p. 548bl; David Parker, pp. 225, 255; Robert Prezioso, pp. 603; Print Collector, p. 129; Science Photo Library, pp. 187, 206, 226cl, 226cr, 232l, 235, 427l, 532tl, 549tr, 564, 583t, 583c; Science Source, p. 205; Dr Gary Settles, p. 438; Staff, p. 290; Gianni Tortoli, p. 221t; Trevor Clifford Photography, pp. 98, 106.

Government of Tasmania, Department of Education: Text adapted from Sustainability Learning Centre – Information Sheet (TASED-4-2058), p. 67.

Ray Joyce/Morrison & Breytenbach Architects: p. 67.

Drs Susan M Landau and William Jagust, Univ of CA, Berkeley: p. 571c.

NASA: pp. 172c, 172cr, 340; Ken Crawford, p. 170; NASA/CXC/JPL-Caltech/STScl, p. 453; NASA, ESA, the Hubble Heritage Team (STScl/AURA), A. Nota (ESA/STScl), and the Westerlund 2 Science Team, pp. 167, 251-4; NASA and WMAP Science Team, p. 180; STSCI/NASA/courtesy of nasaimages.org, p. 482; J.R. Petit et al/Courtesy NASA/JPL-Caltech, p. 43.

Newspix: Jon Hargest, p. 293.

Pearson Education: Alice McBroom, p. 656b; HL Studios. Pearson Education Ltd, p. 535c; Oxford Designers & Illustrators Ltd, pp. 539b, 545.

Wilhelm Röntgen: p. 554.

Science Photo Library: pp. 322, 459b, 560; Steve Allen, p. 535tr; Amelie-Benoist/BSIP, p. 216; American Institute of Physics, p. 226bl; Carl Anderson, p. 196cl; Arscimed, p. 190; Atlas Collaboration, p. 196b; Alex Bartel, p. 155; John Bavosi, p. 512; Martin Bond, p. 77; H.E. Bond/E. Nelan/M. Barstow/M. Burleigh/J.B. Holberg/NASA/ESA/StSci, p. 470; Dr John Brackenbury, p. 611l; Massimo Brega,The Lighthouse, p. 546; BSIP, Raguet, p. 557t; CC Studio, p. 574tr; CERN, pp. 585t, 593t, 596; Chris Butler, p. 458cr; Martyn F. Chillmaid, p. 518; I. Curie & F. Joliot, p. 195; Department of Physics, Imperial College, pp. 245cl; 245bl; Luke Dodd, p. 469cr; Du Cane Medical Imaging Ltd, p. 574tl; Patrick Dumas, p. 153; Equinox Graphics, p. 186; European Southern Observatory, p. 478; Simon Fraser, p. 569t; Mark Garlick, pp. 483br, 540r; Robert Gendler, p. 454; GustoImages, p. 103; Tony & Daphne Hallas, p. 172tl; David A. Hardy, p. 42; Adam Hart-Davis, p. 623; Phil Hill, p. 548c; Gary Hincks, p. 47; Hipersynteza, p. 598t; IBM Research, p. 599r; Mikkel Juul Jensen, p. 38; Manfred Kage, p. 243; Dorling Kindersley/UIG, pp. 619tr, 457br; James King-Holmes, pp. 590br, 593cr; Mehau Kulyk, p. 488; Laguna Design, p. 583b; Lawrence Berkeley Laboratory, pp. 194; 585b; Lawrence Livermore National Laboratory, p. 234; Jon Lomberg, p. 469tl; Davide de Martin, p. 3; Dr P. Marzzi, p. 574b; Thomas McCauley, Lucas Taylor/CERN, p. 590tl; John Mead, p. 549b; Medical Photo NHS Lothian, p. 524; NASA, pp. 303b, 321b; NASA/CXC/SAO, p. 479c; NASA/CXC/University of Wisconsin, p. 459t; NASA/ESA/Hubble Heritage Team (STSCI/AURA), p. 484; NASA/ESA/STSCI/J. Hester & A. Loll, ASU, p. 479b; NASA/ESA/STSCI/J. Hester & P. Scowen, ASU, p. 476b; NASA/ESA/STSCI/R. Windhorst & S. Pascarelle, ASU, p. 184; NASA/Goddard Space Flight Center Scientific Visualization, p. 55; NASA/JPL, p. 35; NASA/SDO/Goddard Space Flight Center, p. 465; NASA/WMAP Science Team, p. 483t; NOAO, p. 473br; National Institute of Health, p. 569b; Ria Novosti, pp. 56, 210t; Ouellettte & Theroux, Publiphoto Diffusion, p. 571br; David Parker, pp. 211, 591, 599l, 599c; Jose Antonio Peñas, p. 473t; George Post, p. 457l; Public Health England, p. 210b, 570b; Ramon Andrade 3dciencia, p. 598cl; Detlev van Ravenswaay, pp. 178, 466; Royal Obersavatory, Edinburgh/AAO, p. 476t; Dr Gary Settles, pp. 605t; 611tr; Alan Sirulnikoff, p. 653; Eckhard Slawik, p. 471; Sheila Terry, p. 619l; Trevor Clifford Photography, pp. 138; 656t; US Department of Energy, pp. 226t, 539t; Tom Van Sant, p. 51; Carol & Mike Werner, p. 40tl; Jim West, p. 76; Loren Winters/Visuals Unlimited, p. 321t.

Scripps Institution of Oceanography, UC San Diego: p. 59.

Shutterstock: AGorohov, pp. 377t, 377c, 377b, 378tl, 378tr, 378bl, 378br, 379l, 379r; Shelby Allison, p. 413r; Galyna Andrushko, p. 332tc; Gurgen Bakhshetsyan, p. 567t; Vitalii Bashkatov, p. 411; bikeriderlondon, p. 273; Joggie Botma, p. 336; Michel Cecconi, p.20; Jaruek Chairak, p. 433b; Ivan Cholakov, p. 295; CLS Design, p. 622tr; Neale Cousland, p. 616; FlashStudio, p. 604t; Hank Frentz, p. 570t; angelo gilardelli, p. 348; Stephen Girimont, p. 338b; Mitch Gunn, p. 617t; Timothy Hodgkinson, p. 143; Iculig, p. 648; jdwfoto, p. 408tr; Dmitry Kalinovsky, pp. 383, 385t; kamilpetran, p. 23; Sebastian Kaulitzki, p. 620l; Milos Kontic, p. 637; KoQ Creative, p. 620br, 622tl, 622bl; Kostsov, p. 408cr; Joana Lopes, p. 665; magnetix, p. 540l; Mart, p. 93r; Mutation, p. 291t; Maxisport, p. 494; Iaroslav Neliubov, p. 332br; nuwatphoto, p. 410; oliveromg, pp. 388t, 388b; ommaphat chotirat, p. 396; Pavel L Photo and Video, p. 556; Photobac, p. 429; Plati Photography, p. 672; Claire Plumridge, p. 376t; Olga Popova, p. 455; Lee Prince, p. 337; Stephen Rees, p. 517bl; Jaco van Rensburg, p. 617b; Jamie Roach, p. 358l; Yuriy Rudyy, p. 611br; Thorsten Schier, p. 544r; Ken Schulze, p. 490; Joao Seabra, p. 607r; Roman Striga, p. 489; Surakit, pp. 380t, 382, 384; Thieury, p. 372; Thirteen, p. 332tr; violetkaipa, p. 91; zstock, p. 64.

Victorian Curriculum and Assessment Authority (VCAA): Selected extracts from the VCE Physics Study Design (2016-2021) are copyright Victorian Curriculum and Assessment Authority (VCAA), reproduced by permission. VCE® is a registered trademark of the VCAA. The VCAA does not endorse this product and makes no warranties regarding the correctness or accuracy of its content. To the extent permitted by law, the VCAA excludes all liability for any loss or damage suffered or incurred as a result of accessing, using or relying on the content. Current VCE Study Designs and related content can be accessed directly at www.vcaa.vic.edu.au, pp. x, 1, 29. 87, 123, 167, 199, 225, 256-7, 287, 331, 375, 411, 453, 489, 531, 553, 577, 603, 645-7.