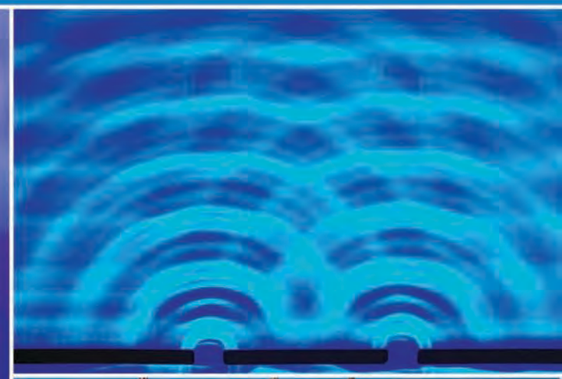
fetena.net : Ethiopian online exams and textbooks

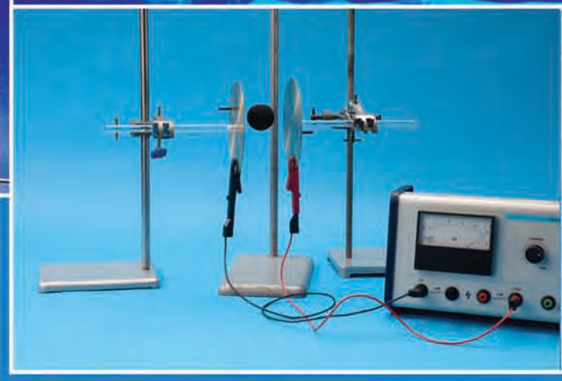
Physics

Student Textbook
Grade 12

Physics Student Textbook Grade 12







Federal Democratic Republic of Ethiopia
Ministry of Education

ISBN 978-99944-2-022-3

Price: ETB 36.00

FDRE
MoE

Federal Democratic Republic of Ethiopia
Ministry of Education



Physics

Student Textbook
Grade 12

Authors: Graham Bone
Tim Greenway

Advisers: Tilahun Tesfaye Deressu (PhD)
Endeshaw Bekele Buli

Evaluators: Yosef Mihiret
Gebremeskel Gebreegziabher
Yusuf Mohamed



Federal Democratic Republic of Ethiopia
Ministry of Education

PEARSON

Acknowledgments

The development, printing and distribution of this student textbook has been funded through the General Education Quality Improvement Project (GEQIP), which aims to improve the quality of education for Grades 1–12 students in government schools throughout Ethiopia.

The Federal Democratic Republic of Ethiopia received funding for GEQIP through credit/financing from the International Development Associations (IDA), the Fast Track Initiative Catalytic Fund (FTI CF) and other development partners – Finland, Italian Development Cooperation, the Netherlands and UK aid from the Department for International Development (DFID).
The Ministry of Education wishes to thank the many individuals, groups and other bodies involved – directly and indirectly – in publishing the textbook and accompanying teacher guide.

The publisher would like to thank the following for their kind permission to reproduce their photographs:

(Key: b-bottom; c-centre; l-left; r-right; t-top)

Alamy Images: 2tl, 13c, 13cl, 37r, 38t, 46cl, 53br, 56cl, 70cl, 73cr, 80bl, 92tl, 100tl, 102tl, 104cl, 110tl, 111c (Circular waves), 111cr (Plane waves), 114tl, 114cl, 115c, 115cl (water waves), 117br, 118tl, 125cr, 135tr, 135br, 144tl, 285t; **Rex Features:** 72bl;
Science Photo Library Ltd: 6bl, 111br, 118cl, 118bl, 121tr, 121br, 123br, 124tl, 147br, 153tr, 191c, 317cl, 324bl;

Cover images: Front: **Science Photo Library Ltd:** cl/Ice, waves;
All other images © Pearson Education

Every effort has been made to trace the copyright holders and we apologise in advance for any unintentional omissions. We would be pleased to insert the appropriate acknowledgement in any subsequent edition of this publication.

© Federal Democratic Republic of Ethiopia, Ministry of Education
First edition, 2002 (E.C.)
ISBN: 978-99944-2-022-3

Developed, Printed and distributed for the Federal Democratic Republic of Ethiopia, Ministry of Education by:
Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

In collaboration with
Shama Books
P.O. Box 15
Addis Ababa
Ethiopia

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the prior written permission of the copyright owner or a licence permitting restricted copying in Ethiopia by the Federal Democratic Republic of Ethiopia, Federal Negarit Gazeta, *Proclamation No. 410/2004 Copyright and Neighboring Rights Protection Proclamation, 10th year, No. 55, Addis Ababa, 19 July 2004.*

Disclaimer
Every effort has been made to trace the copyright owners of material used in this document. We apologise in advance for any unintentional omissions. We would be pleased to insert the appropriate acknowledgement
in any future edition

Printed in Malaysia, CTP-PJB

Contents

Unit 1 Thermodynamics 1

1.1 Thermal equilibrium and definition of temperature 3
1.2 Work, heat and the first law of thermodynamics 9
1.3 Kinetic theory of gases 21
1.4 Second law of thermodynamics, efficiency and entropy 31
1.5 Heat engines and refrigerators 37

Unit 2 Oscillations and waves 51

2.1 Periodic motion (basic concepts) 53
2.2 Wave motion 80
2.3 Sound, loudness and the human ear 97

Unit 3 Wave optics 108

3.1 Wave fronts and Huygens's principle 109
3.2 Reflection and refraction of plane wave fronts 113
3.3 Proof of the laws of reflection and refraction using Huygens's principle 116
3.4 Interference 120
3.5 Young's double slit experiment and expression for fringe width 128
3.6 Coherent sources and sustained interference of light 131
3.7 Diffraction due to a single slit and a diffraction grating 133

Unit 4 Electrostatics 141

4.1 Electric charge and Coulomb's law 143
4.2 Electric potential 162
4.3 Capacitors and dielectrics 173

Unit 5 Steady electric current and circuit properties 198

5.1 Basic principles 199
5.2 Kirchoff's rules 214
5.3 Measuring instruments 220
5.4 The Wheatstone bridge and the potentiometer 226

Unit 6 Magnetism 234

6.1 Concepts of a magnetic field 235
6.2 The Earth and magnetic fields 238
6.3 Motion of charged particles in a magnetic field 240
6.4 Magnetic force on current-carrying conductors (long, staight, circular loop) 247
6.5 Ampere's law and its application 256
6.6 Earth's magnetism 260

Unit 7 Electromagnetic induction and a.c. circuits 266

7.1 Phenomena of electromagnetic induction 268
7.2 Alternating current (a.c.) generator and transformers 281
7.3 Alternating current (a.c.) 287
7.4 Power in a.c. circuits 304

Unit 8 Atomic physics 311

8.1 Dual nature of matter and radiation 312
8.2 Atoms and nuclei 322

Index 348

Thermodynamics

Unit 1

Contents

Section	Learning competencies
1.1 Thermal equilibrium and definition of temperature (page 3)	• Define the terms atomic mass, mole, molar mass and Avogadro's number. Use their relationship to solve related problems. • Define the zeroth law of thermodynamics. • Determine the relationship between temperature and energy transfer and thermal equilibrium. • State what is meant by an absolute scale of temperature. • Draw phase diagrams to determine the triple point of a substance. • Differentiate between the critical point and the boiling point of a substance.
1.2 Work, heat and the first law of thermodynamics (page 9)	• Give the definitions of isothermal, isobaric, isochoric and adiabatic processes and draw their associated p–V diagrams. • Calculate work and heat for ideal gas processes. • State the first law of thermodynamics. • Identify the appropriate form of the first law of thermodynamics for isobaric, isochoric and isothermal processes. • Describe ways of changing the internal energy of a gas. • Apply thermodynamics laws to solve simple numerical problems. • Solve problems involving calculations of pressure, temperature or volume for a gas undergoing adiabatic changes. • Define molar heat capacity. • Distinguish between molar heat capacity at constant pressure and at constant volume. Show their relationship based on Mayer's equation. • Show that the molar heat capacity at constant pressure is greater than the molar heat capacity at constant volume. • Evaluate $C_p - C_v$ and $\frac{C_p}{C_v}$ for an ideal gas. • Indentify the value of $\frac{C_p}{C_v}$ for atomic gases and monatomic gasses. • Use $TV^{\gamma-1}$ = constant for adiabatic processes to solve problems.
1.3 Kinetic theory of gases (page 21)	• State the assumptions made to define an ideal gas. • Describe the kinetic theory of gases, including the importance of Brownian motion and diffusion. • Define r.m.s. velocity of a gas and the mean free path for a gas particle. • Use the expression for the pressure of an ideal gas in terms of its density and mean square speed of molecules to solve problems. • Solve problems to determine p, V, T or r.m.s. speed of gas molecules for an ideal gas, given relevant data. • State Graham's law of diffusion and use it to solve related problems. • State Dalton's law of partial pressure and use it to solve related problems.

Contents

Section	Learning competencies
1.4 Second law of thermodynamics, efficiency and entropy (page 31)	• State the second law of thermodynamics. • Appreciate that the second law of thermodynamics places sharp constraints on the maximum possible efficiency of heat engines and refrigerators. • Distinguish between reversible and irreversible processes. • Define entropy as a measure of disorder and state the second law of thermodynamics in terms of entropy.
1.5 Heat engines and refrigerators (page 37)	• Describe the fundamental principles of heat engines and refrigerators. • Solve problems involving heat flow, work and efficiency in a heat engine. • Identify that all real heat engines lose some heat to their surrondings. • Investigate the physical principles that all heat engines and refrigerators must obey.



Figure 1.1 *Sadi Carnot was born in Paris in 1796. He was the first person to start thinking about how to better the steam engines that were proving so valuable during the Industrial Revolution.*

KEY WORDS

thermodynamics *the study of energy conversion between heat and other forms of energy*

Newton's laws *laws describing the relationship between the forces acting on a body and its motion due to those forces*

Another title for this unit could be "The universe and everything in it". Thermodynamics is a grand topic. Historically, it has humble roots in the Industrial Revolution of the 19th century and the rise of the steam engine as a way of harnessing the power of nature to meet the needs of mankind. The term is derived from the Greek words *therme*, meaning heat, and *dynamis*, meaning power. Following the subsequent discovery of the atom, this new branch of physics has grown to help us now understand the interaction between energy and matter on a multitude of spatial scales: from the interaction between individual atoms, through all processes and events that occur in our own everyday surroundings, to stars, galaxies and even the universe as a whole.

As physicists, our role is often to take a reductionist approach: that is, to reduce the world around us to more simple, more fundamental situations that allow us to use the basic laws we have at our disposal to solve specific problems. **Thermodynamics** is different. For example, we might wish to work out the effect of heating a gas in a sealed container. Alternatively, we might squash the gas. We could work out that the pressure of the gas on the walls of the container would increase in both cases by applying the laws of Newtonian mechanics to every molecule in the gas. **Newton's laws** are very simple but this would be an extremely time-consuming process. Instead, the laws of thermodynamics provide us with solutions to these problems whilst requiring knowledge of only large-scale quantities of the gas such as volume, temperature and pressure.

As they do not depend on the specific details of the systems being studied, the laws of thermodynamics are extremely powerful tools that allow us to predict the behaviour of systems like machines, engines and, in particular, gases. These same laws also help to provide us with answers to deep questions like:

- Why does the "arrow of time" only ever point in one direction?
- How will the universe end?
- What will happen to an object if you keep cooling it down?

1.1 Thermal equilibrium and definition of temperature

By the end of this section you should be able to:

- Define the terms atomic mass, mole, molar mass and Avogadro's number. Use their relationship to solve related problems.
- Define the zeroth law of thermodynamics.
- Determine the relationship between temperature and energy transfer and thermal equilibrium.
- State what is meant by an absolute scale of temperature.
- Draw phase diagrams to determine the triple point of a substance.
- Differentiate between the critical point and the boiling point of a substance.

Before we can learn the laws of thermodynamics and be able to apply them successfully, we must first be confident of our understanding of the particle model of matter, and terms such as heating and thermal equilibrium. Temperature is more difficult to define and we will encounter a number of different ways to approach temperature in this unit.

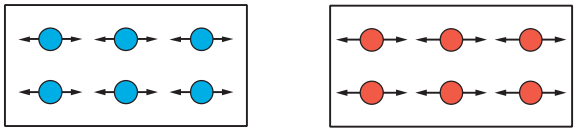


Figure 1.2 *Temperature is related to the average random kinetic energy of particles in a substance.*

Atoms and moles

In order to understand temperature and heat at the particle level we need to use and understand the terms moles and molar mass.

The mole (unit, mol) is a measure of number of discrete particles in a substance (solid, liquid or gas). From definition:

- 1 mol = 6.02×10^{23} particles

1 mol of helium gas contains 6.02×10^{23} particles of helium. However care must be taken when dealing with more complex substances. Each water molecule contains two hydrogen atoms and one oxygen atom (H_2O). 1 mol of water contains 6.02×10^{23} molecules of water. This equates to 6.02×10^{23} atoms of oxygen and 1.20×10^{24} atoms of hydrogen (as there are two hydrogen atoms in each water molecule).

DID YOU KNOW?

6.02×10^{23} is called Avogadro's number (N_A); it is defined as the number of particles in 12 g of carbon-12.

The relationship between the number of moles, the number of particles and Avogadro's number is shown below:

- $N = nN_A$

where

N = number of particles

n = number of moles

N_A = Avogadro's number, 6.02×10^{23} mol^{-1}

KEY WORDS

molar mass *the mass of one mole of a substance*

kinetic energy *the energy possessed by an object as a result of its motion*

thermal energy *the energy possessed by an object resulting from the movement of its particles*

internal energy *the sum of the random distribution of kinetic and potential energies associated with the molecules within a system*

zeroth law *two bodies that are separately in thermal equilibrium with a third body must be in thermal equilibrium with each other*

thermal equilibrium *condition in which two bodies are at the same temperature and there is no net transfer of energy between them*

Discussion activity

It is often helpful to know the mass of 1 mol of a substance. This is called the molar mass. One mole of carbon-12 has a **molar mass** of 12 g or 0.012 kg. Uranium-238 has a molar mass of 238 g or 0.238 kg. If you know the number of moles of a substance and the molar mass of you can calculate the mass of the substance.

- $m = nM$

where

m = mass of substance

n = number of moles

M = molar mass

This equation can be combined with the previous one to give:

- $n = \frac{N}{N_A} = \frac{m}{M}$

From this the mass of each particle (m_p) can be calculated using:

- $m_p = \frac{M}{N_A}$

or

- $m_p = \frac{m}{N}$

Worked example 1.1

A block of pure carbon-12 contains 4.2 mol particles. Calculate:

a) the number of particles in the block

b) the mass of the block

c) the mass of each carbon-12 atom.

a)
- $N = nN_A$ — *Use the relationship between number of moles and Avogadro's number*
- $N = 4.2 \times 6.02 \times 10^{23}$ — *Substitute the known values*
- $N = 2.5 \times 10^{24}$ particles — *Solve the equation*

b)
- $m = nM$ — *Use the relationship between number of moles and Molar mass*
- $m = 4.2 \times 0.012$ — *Substitute the known values*
- $m = 0.05$ kg — *Solve the equation*

c)
- $m_p = \frac{M}{N_A}$ — *Use the relationship between number of moles and Molar mass*
- $m_p = \frac{0.012}{6.02 \times 10^{23}}$ — *Substitute the known values*
- $m_p = 2.0 \times 10^{-26}$ kg — *Solve the equation*

What causes thermal energy transfer?

The particle model of matter is a way of visualising what individual particles are doing inside a substance. In solids, particles are vibrating. In liquids and gases, the particles are moving more freely. In all cases, the particles have kinetic energy; this causes them to collide with their neighbours. As a result, energy is distributed throughout the substance. There is a range of different particle speeds and directions within the substance: so many, in fact, that we say that the motion is random. However, on average if the particles in the substance are vibrating faster we would say that the substance is hotter – the temperature is higher.

Temperature is something that we all have experience of. If we place two bodies of different temperatures in contact, then the particles at the boundary will collide and the kinetic energy of particles is transferred backwards and forwards between the objects. A 'body' is another word for an object. On average, the particles in the hotter body have more **kinetic energy** than those in the colder body, so there is a net transfer of **thermal energy** from the hotter body to the colder body. This process is referred to as heating. This is the only way that the word heat can be used. A body does not contain or possess heat. This is just the same as an electrical component, which does not contain or possess electrical current. Instead we will use the term **internal energy** to describe the total energy that is internal to bodies.

- Temperature is a measure of the average random kinetic energy of particles in a body, and is used to determine in which direction there will be a net energy flow when two bodies are close to one another.

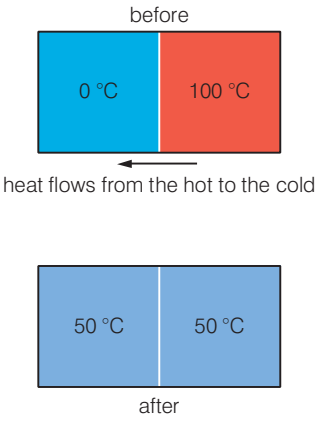


Figure 1.3 *Heat flows from the hotter body to the colder body until the temperatures are equal.*

Activity 1.1: Hot and cold water

Place one hand in a bucket of cold water and the other hand in a bucket of warm water. Think about what is happening to the particles in your hand and in the water. In which direction is heat transfer taking place and why?

DID YOU KNOW?

Historically, of the four laws of thermodynamics, the first to be derived was the so-called **second law**. The **first law** came next. The **third law** was only formulated in 1912 and the **zeroth law**, although very important, was developed as something of an afterthought.

What is the zeroth law?

The zeroth law of thermodynamics states that:

"Two bodies that are separately in thermal equilibrium with a third body must be in thermal equilibrium with each other."

When two bodies are in **thermal equilibrium** then there is no net transfer of energy between them.

From our everyday experience, the zeroth law may seem obvious, but it provides us with a way of defining temperature: it is the property of a body that determines whether it is in thermal equilibrium with other bodies. This also enables accurate calibration between thermometers of different kinds.

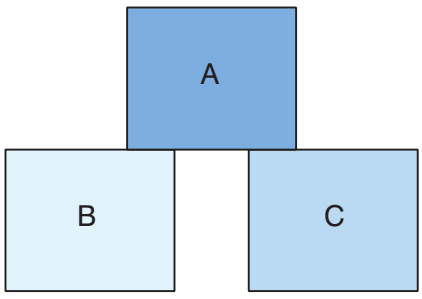


Figure 1.4 *If A is in thermal equilibrium with B, and C is in thermal equilibrium with B, then A is also in thermal equilibrium with C.*

DID YOU KNOW?

The term **steady state** can be confused with the term thermal equilibrium. Steady state indicates that a situation is not changing with time. There may be a constant temperature difference between two bodies so there may be a steady non-zero net flow of energy between them. Thermal equilibrium implies steady state, but the reverse is not true.

KEY WORDS

steady state *a situation that is not changing with time*

DID YOU KNOW?

The Kelvin temperature scale is referred to as 'absolute' as it is independent of any other property of a substance.



Figure 1.6 *The triple point of water is the temperature (and pressure) at which all three phases (ice, water and water vapour) coexist in thermal equilibrium.*

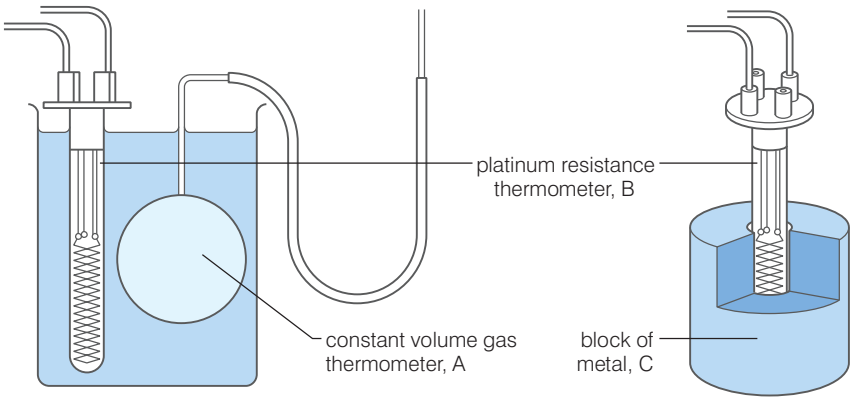


Figure 1.5 *The constant volume gas thermometer, A, indicates a consistent water temperature T and the platinum resistance thermometer, B, indicates a consistent water temperature of $T + \Delta T$. When the platinum resistance thermometer is then placed in contact with the block of metal it indicates a consistent temperature of $T + \Delta T$. The water, the block of metal and the two thermometers are therefore all in thermal equilibrium.*

The thermodynamic temperature scale

The accurate and reliable measurement of temperature has always posed problems for scientists. An international conference in 1947 decided that the absolute temperature scale that should be used was the thermodynamic, or ideal gas, temperature scale measured in kelvin (K). **Absolute zero**, the temperature at which all the random motion of particles in a substance ceases, is defined as 0 K and the triple point of water, the temperature at which all three states of water can coexist, is defined as 273.16 K. In relation to the Celsius temperature scale, t, the **thermodynamic scale**, T, is then given by:

$$T(\text{K}) = t(^\circ\text{C}) + 273.15$$

	Absolute temperature (K)	Celsius temperature (°C)
Absolute zero	0.00	–273.15
Triple point of water	273.16	0.01
Ice point	273.15	0.00
Steam point	373.15	100.00
Room temperature	293	20

Table 1.1 *Important temperatures in units of kelvin and degrees Celsius.*

Activity 1.2: Celsius and kelvin

Research the temperature of a range of objects (e.g. the surface of the Sun, a healthy human body temperature, etc.) and express these in °C and in K.

Discussion activity

The third law of thermodynamics is not a law in the same way as the others. But for completeness, it states: "no object can reach a temperature of **absolute zero** in a finite number of steps." Can you explain why this is true?

KEY WORDS

absolute zero *the temperature at which all random motion of particles in a substance ceases*

thermodynamic scale *an absolute measurement of temperature*

phase (of matter) *point at which all the physical properties within a material are uniform*

phase diagram *graph of pressure against temperature for a given substance*

Phases of matter

A **phase** of matter is when all the physical properties within a material are uniform. At a given phase of matter a material will have the same density and refractive index.

Take the example of some ice cubes in a glass of water. The ice cubes are at one phase, the water is at another and there is a third phase just above the water as it evaporates.

States of matter are classifications of the distinct phases of matter based on their large-scale properties. For example, a solid phase is the state in which the substance maintains a fixed volume and a fixed shape, whereas the liquid phase is one where the substance can change to take the shape of its container.

Changing the pressure or temperature of a substance will affect its phase. Plotting a graph of pressure against temperature is often called a **phase diagram**.

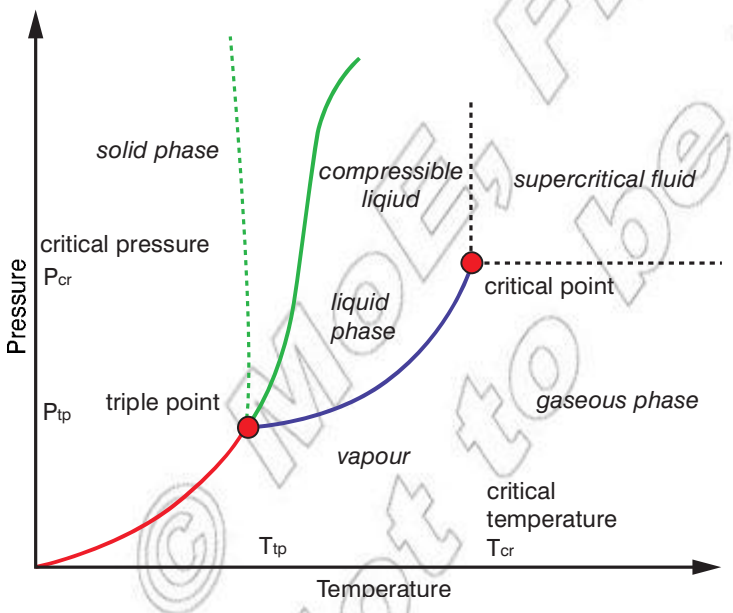


Figure 1.7 A phase diagram showing different states of matter

From the graph there are clear boundaries between different phases and two points of particular interest.

Triple point

As we've already mentioned this is the specific temperature and pressure where all three states of matter are able to exist in thermal equilibrium.

© MoE FDR Ethiopia Not to be republished

Critical point (sometimes called a critical state)

The critical point occurs where the critical temperature and critical pressure of a substance meet. Above this point clear phase boundaries cease to exist. For water the critical point is around 647 K and 22 MPa. The properties of the gas and liquid phases merge together giving only one phase at the critical point. The substance forms a supercritical fluid. Above the critical temperature it is not possible to form a liquid, regardless of any increase in pressure.

The critical point is different to the boiling point of a substance. The boiling point of a substance is usually meant to mean the boiling point at standard atmospheric pressure. However, in Figure 1.7 the actual boiling point of a substance depends on the surrounding pressure. From the phase diagram the substance has a number of boiling points following the curving blue line. The lower the pressure, the lower the boiling point.

The critical point only occurs at one specific temperature and pressure.

Summary

- 1 mol = 6.02×10^{23} particles
- The zeroth law of thermodynamics: two bodies that are separately in thermal equilibrium with a third body must be in thermal equilibrium with each other.
- Temperature defines the direction of net energy transfer between two bodies.
- When two bodies are in thermal equilibrium then there is no net transfer of energy between them and they are both at the same temperature.
- The absolute temperature scale: $T(\text{K}) = t(°\text{C}) + 273.15$

Review questions

1. Convert the following to degrees Celsius:
   a) the boiling point of helium, 4.25 K
   b) the freezing point of gold, 1340 K.
2. Convert the following to kelvin:
   a) the freezing point of mercury, –39 °C
   b) the average temperature of the universe, –270.42 °C.
3. In Figure 1.5, why was it necessary to say that the readings on the thermometers were consistent?
4. Sketch a phase diagram labelling and explaining any key points.

© MoE - Not to be republished

1.2 Work, heat and the first law of thermodynamics

By the end of this section you should be able to:

- Give the definitions of isothermal, isobaric, isochoric and adiabatic processes and draw their associated p–V diagrams.
- Calculate work and heat for ideal gas processes.
- State the first law of thermodynamics.
- Identify the appropriate form of the first law of thermodynamics for isobaric, isochoric and isothermal processes.
- Describe ways of changing the internal energy of a gas.
- Apply thermodynamics laws to solve simple numerical problems.
- Solve problems involving calculations of pressure, temperature or volume for a gas undergoing adiabatic changes.
- Define molar heat capacity.
- Distinguish between molar heat capacity at constant pressure and at constant volume. Show their relationship based on Mayer's equation.
- Show that the molar heat capacity at constant pressure is greater than the molar heat capacity at constant volume.
- Evaluate $C_p - C_v$ and $\frac{C_p}{C_v}$ for an ideal gas.
- Identify the value of $\frac{C_p}{C_v}$ for atomic gases and monatomic gasses.
- Use $TV^{\gamma-1}$ = constant for adiabatic processes to solve problems.

KEY WORDS

caloric *a 19th century representation of heat as an invisible fluid*

In the early 19th century, steam engines were becoming more popular and were replacing traditional methods of doing work, for example, moving objects and people from one place to another. Sadi Carnot and many other scientistists of the time believed that heat was like an invisible, massless fluid – it was then called '**caloric**'. The steam engine, according to Carnot, worked like a water mill with caloric running 'downhill' from the boiler to the condenser and, in the process, driving the shafts of the engine.

Twenty years later, James Prescott Joule, working in his father's brewery in England, showed that the temperature of a body could be increased just by doing mechanical work on it. He called this process the 'mechanical equivalent of heat'. His work showed that the idea of heat as a fluid running 'downhill' was false. He also paved the way for a greater understanding of energy and temperature leading to the first law of thermodynamics.

© Ethiopia Not to be published

KEY WORDS

Carnot's principle *a principle that sets a limit on the maximum efficiency any possible engine can obtain*

system *a body that is the subject of study and that may be solid, liquid or gas*

surroundings *the environment that is around a system and in thermal contact with it*

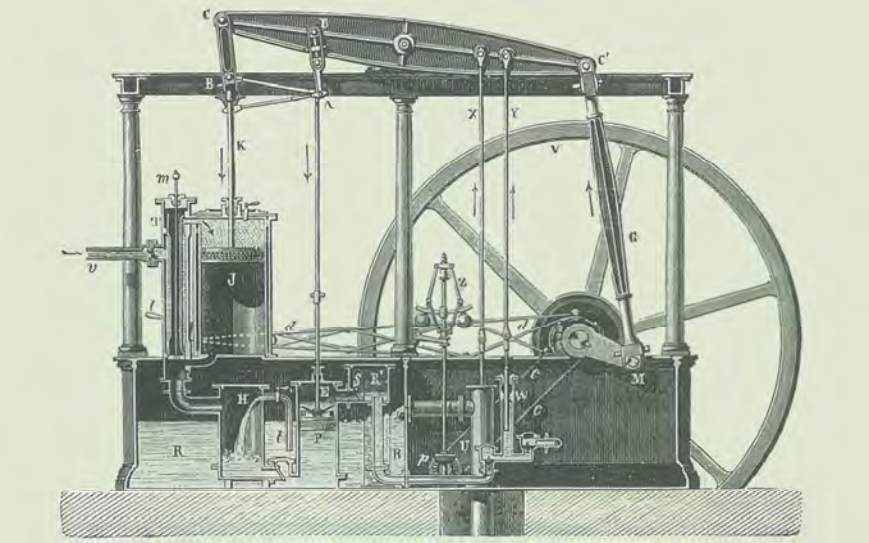


Figure 1.8 *An example of an early steam engine. Just like the amount of water stays constant as it flows through a water mill, so Carnot thought that the total amount of caloric was unchanged as it ran through a steam engine. This part of his idea was later shown to be incorrect but* ***Carnot's principle*** *is still used today to determine the maximum efficiencies of boilers and turbines in power stations.*

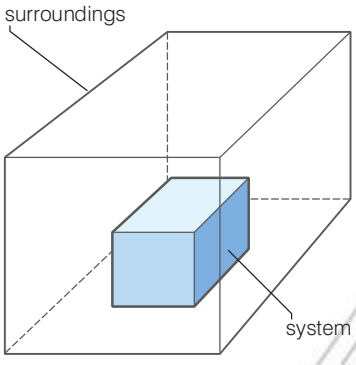


Figure 1.9 *A representative diagram of a system and its surroundings*

What exactly is internal energy?

In the topic of thermodynamics we always study a **system** in relation to its **surroundings**. The system is another word for a body that may be solid, liquid or gas, complex or simple. It is the 'thing' that we are interested in. The surroundings might be the table an object is resting on and the atmosphere around it, or it may be the entire universe. It is that which surrounds the system and is in thermal contact with it.

It seems common sense that if we heat a system then that system will become hotter. But is this always true? What if the system was a small beaker of water at room temperature placed above a candle flame? Its temperature will increase until it reaches 100 °C, but the temperature will then stay constant as the water boils. Clearly we have transferred energy to the water at a constant rate. At the start of the heating process the energy transferred increased the average random kinetic energy of the molecules in the water. But during the phase change from liquid to gas, work was done on individual molecules to move them apart against their intermolecular attraction. The internal energy of the steam at 100 °C is greater than that of the same mass of water at 100 °C.

- The internal energy of a system is the sum of the random distribution of kinetic and potential energies associated with the molecules within that system.

© MoE Not to be republished

Discussion activity

What about ideal gases? Within an ideal gas, there are no intermolecular forces except during collisions, so the internal energy is equal to the sum of random kinetic energies of the particles only.

What about melting ice? When ice melts, the water molecules in the liquid phase are closer to each other, on average, than the molecules in the solid phase. However, energy is still required in order to break the hydrogen bonds that exist between the molecules in ice, and so the internal energy still increases.

What is the first law of thermodynamics?

The work of Joule mentioned at the start of this section led to the idea that energy as a quantity is conserved whenever any process takes place. This notion is expressed most often as the 'law of conservation of energy', which is a simplification of **the first law of thermodynamics**. The first law states that:

"The increase in internal energy of a system is equal to the sum of the energy entering the system through heating, and the work done on the system."

When defining the three quantities, particular attention must be paid to the sign of each quantity. These have the following definitions:

ΔU = *increase* in internal energy of the system

ΔQ = the amount of energy transferred to the system by heating it (that is, by means of a temperature gradient)

ΔW = the amount of work done on the system

The first law of thermodynamics is therefore written as:
$\Delta U = \Delta Q + \Delta W$

KEY WORDS

the first law of thermodynamics *the increase in internal energy of a system is equal to the sum of the energy entering a system through heating, and the work done on the system*

adiabatic *a process that involves no energy transfer into or out of a system as a result of heating*

The bicycle pump

An **adiabatic** process involves no energy transfer into or out of a system as a result of heating. For example, if we push the plunger of a bicycle pump very rapidly whilst blocking the hole at the end of the pump with our finger, the air inside the pump is compressed. The mechanical work done *on* the air results in an *increase* in the internal energy of the air. As there is no phase change, the average potential energy of the molecules stays constant but the average random kinetic energy of the molecules increases; evidence of this is an increase in the temperature of the air. As the compression takes place rapidly, there is a negligible time in which energy transfer to the surroundings by heating can occur, and the process is nearly adiabatic. Of course, after the compression is complete, the temperature gradient between the system and its surroundings will result in an energy transfer *from* the system, resulting in a *decrease*

© Not for Ethiopia
Not to be republished

in the internal energy of the air (and the warming of the pump, the surrounding air and your hand).

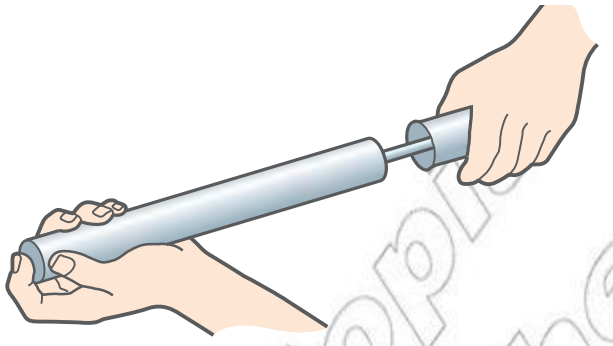


Figure 1.10 *Compressing the air in a bicycle pump*

KEY WORDS

isothermal *a process which involves no change in the temperature of the system*

An **isothermal** process involves no change in the temperature of the system. In these types of processes heating of the system or by the system must be allowed to occur. In practice, this means that the process must occur extremely slowly, also called quasistatically, so that a tiny temperature gradient can be set up allowing the flow of heat into or out of the system. If we were to ensure that the bicycle pump were made of metal, to allow the fast transfer of energy out of the system through heating, and we push the plunger in very slowly whilst blocking the hole at the end of the pump with our finger, the compression of the air inside the pump is very nearly isothermal.

- An adiabatic process involves no transfer of energy into or out of the system as a result of a temperature gradient.
- An isothermal process involves no change in the temperature of the system.

How can we calculate the increase in the internal energy of a system?

1. *Heating the system*. For example, if we transfer energy to a volume of water in a beaker by heating the beaker using a Bunsen flame. If no work is done on or by the system then $\Delta W = 0$ and:

$$\Delta U = \Delta Q = mc\Delta T$$

where m is the mass of the water, c is the specific heat capacity of the water and ΔT is the *increase* in the temperature of the water.

2. *Doing electrical work on the system*. For example, if we pass an electric current, I, through a wire by placing a potential difference, V, across the wire for a time, Δt. If the process was adiabatic then $\Delta Q = 0$ and:

$$\Delta U = \Delta W = VI\Delta t$$

3. *Doing mechanical work on the system*. For example, if we depress a frictionless plunger in a sealed syringe containing helium gas by a distance, Δx, using a constant force, F. If the process was adiabatic then $\Delta Q = 0$ and:

$$\Delta U = \Delta W = F\Delta x$$

© MoE-FDF-Employed Not to be republished

Alternatively, if we know that the volume of the helium was changed by an amount ΔV, whilst the pressure on the gas remained at a constant value, p, then we can calculate the change in internal energy using:

$$\Delta U = \Delta W = -p\Delta V$$

The negative sign is necessary as a decrease in volume of the gas results in work being done on the gas and a subsequent increase in internal energy.

Discussion activity

Can you show that the expressions for mechanical work, $F\Delta x$ and $p\Delta V$, are equivalent?

Remember: doing work on a system is very different to heating it. Working is an ordered process that has nothing to do with a temperature gradient. It does not matter whether the system is hotter or colder than its surroundings. You can still do work on it by applying a force over a certain distance. Heating, on the other hand, is a random process that can only transfer energy down a temperature gradient.

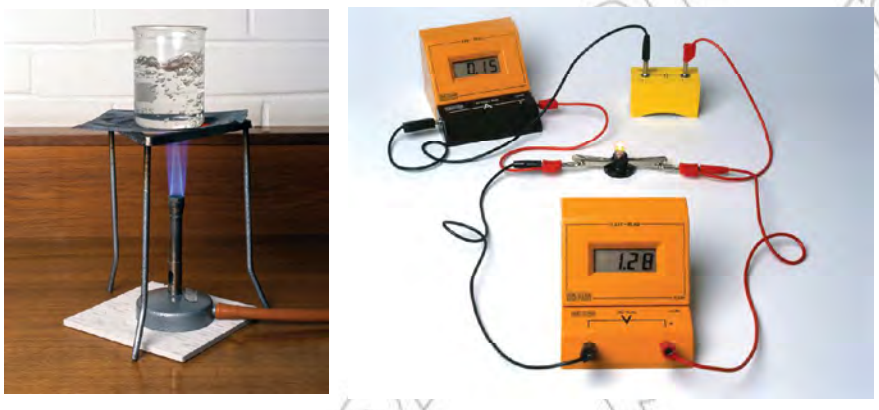


Figure 1.11 *Examples of how to increase the internal energy of a system: (a) heating a volume of water; (b) passing an electric current through a wire inside a bulb*

Discussion activity

Insulated systems allow no transfer of energy by heating. **Isolated** systems do not interact with their surroundings so their internal energy remains constant. Very few objects, including those in space, are truly isolated. Perhaps the only isolated system in existence is the entire universe. If the internal energy of the universe is constant then where did all the energy come from originally?

KEY WORDS

insulated *a system that allows no transfer of energy by heating*

isolated *a system that does not interact with its surroundings, hence its internal energy remains constant*

Not to be republished © Ethiopia

Discussion activity

How is the first law of thermodynamics related to the law of heat exchange that we have studied previously? Which law is more general? As a reminder, the law of heat exchange states that, for an isolated system:

$$\sum Q_{lost} + \sum Q_{gained} = 0$$

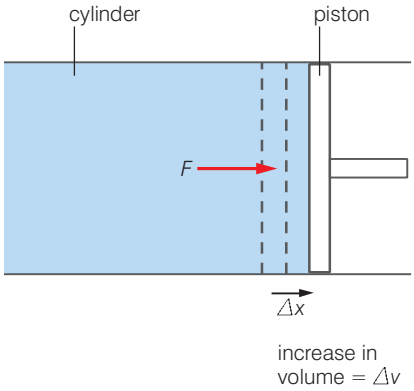


***Figure 1.12** Expansion of ammonia gas in a cylinder*

Example calculation using the first law of thermodynamics

Imagine ammonia gas expanding rapidly within a cylinder forcing a frictionless piston to move a distance of 5.0 cm, as shown in Figure 1.12. The process is isobaric, that is, the pressure of the gas remains constant, at 130 kPa. The mass of the gas is 650 mg and the area of the piston in contact with the gas is 3.0×10^{-3} m^2. What happens to the temperature of the ammonia?

(Specific heat capacity of ammonia at constant pressure = 2.2 kJ/kg K)

The change in volume of the gas

$= 3.0 \times 10^{-3}$ m^2 $\times$ 0.050 m
$= 1.5 \times 10^{-4}$ m^3

The mechanical work done by the gas on the piston

$= p\Delta V$
$= 1.3 \times 10^{5}$ Pa $\times$ 1.5×10^{-4} m^3
$= 19.5$ J

As the process occurs rapidly it is nearly adiabatic, $\Delta Q = 0$ and the first law of thermodynamics reduces to:

$$\Delta U = \Delta W = -p\Delta V$$
$$= -19.5 \text{ J}$$

Since work was done by the gas on its surroundings the internal energy of the gas has decreased. Assuming no change of phase of the gas, we can now find out by how much the temperature of the gas has decreased:

$$\Delta U = mc\Delta T$$
$$-19.5 \text{ J} = 6.5 \times 10^{-4} \text{ kg} \times 2200 \text{ J/kg K} \times \Delta T$$
$$\Delta T = -14 \text{ K}$$

The temperature of the ammonia gas has decreased by 14 K.

© MoE, FDR Eritrea
Not to be republished

Activity 1.3: Temperature of a rubber band

Carry out the following practical where the system being studied is a short piece of rubber band. For each step, write down whether the temperature of the rubber increased or decreased. Also, for each step, write down whether the values for ΔU, ΔQ and ΔW are negative, zero or positive.

1. Place the rubber band against your lower lip and stretch it *quickly*.
2. Keep the rubber band stretched against your lower lip for 10 seconds.
3. Still holding the rubber against your lip, *quickly* allow it to return to its original length, but keep hold of it!
4. Keep the rubber band in contact with your lower lip for 10 seconds.

The first law for different thermodynamic processes

Isobaric

During an **isobaric process** the pressure of the system remains constant. Any work done by the system will result in an increase in volume. For example, imagine a gas expanding at constant pressure.

The yellow area represents the work done. This is given by:

- $\Delta W = p\Delta V$

Substituting this into the first law equation gives:

- $\Delta U = \Delta Q + \Delta W$
- $\Delta U = \Delta Q + p\Delta V$

Remember the $+\Delta W$ represents work done on the system. In the case of the expanding gas, work is being done by the system and the equation would be written as:

- $\Delta U = \Delta Q - p\Delta V$

Isochoric

If the volume of the system remains constant then there is no mechanical work down on the system. Since $\Delta W = p\Delta V$, if $\Delta V = 0$, ΔW must also equal 0. The first law equations for an isochoric process would be written as:

- $\Delta U = \Delta Q + \Delta W$
- $\Delta U = \Delta Q + 0$
- $\Delta U = \Delta Q$

KEY WORDS

isobaric process *a process in which the pressure of a gas remains constant*

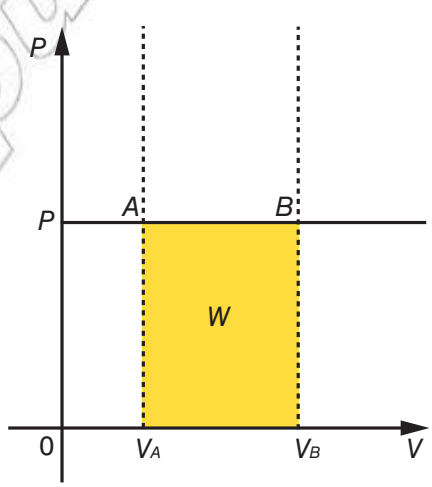


Figure 1.13 *A simple p–V digram for an isobaric process*

© MoE FDR Ethiopia Not to be republished

Activity 1.4 Thermo-dynamic processes

Draw a table showing the different forms of the first law for the four processes described.

Isothermal

If the process occurs at constant temperature then there is no change in the internal energy of the system so $\Delta U = 0$. The first law equations for an isothermal process would be written as:

- $\Delta U = \Delta Q + \Delta W$
- $0 = \Delta Q + \Delta W$
- $\Delta Q = -\Delta W$

Adiabatic

During an adiabatic process there is no heat transfer into or out of the system, therefore $\Delta Q = 0$. The first law equations for an adiabatic process would be written as:

- $\Delta U = \Delta Q + \Delta W$
- $\Delta U = 0 + \Delta W$
- $\Delta U = \Delta W$

What general expressions can we derive for a gas when it is heated?

If the system we are investigating is a gas then pressure, p, and volume, V, are two important properties of that gas and we often represent the processes that change the state of the gas using a p–V graph. Examples of adiabatic, isothermal, isobaric and isochoric processes are provided in Figure 1.14.

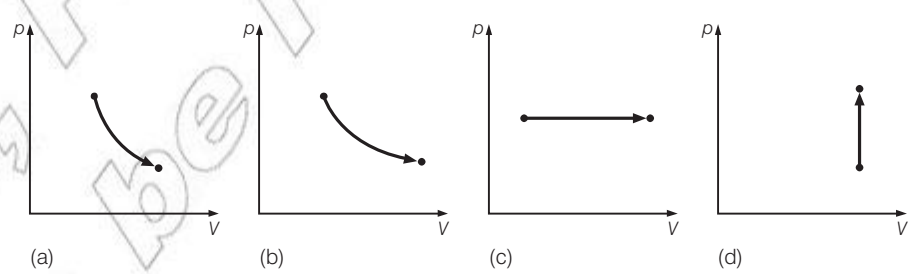


Figure 1.14 *p–V graphs of a gas under (a) adiabatic expansion; (b) isothermal expansion; (c) isobaric expansion; and (d) an isochoric increase in pressure. An isochoric process is one where the volume of the gas is constant.*

KEY WORDS

molar heat capacity *the heat capacity per unit mole of a substance*

In each case the work done is represented by the area under the line. If the gas is doing work (i.e. the arrow is moving to the right) this area should be expressed as $-\Delta W$.

The physics of the gas phase is more accessible than that of the liquid or solid phase. So it is helpful to apply, in detail, the first law of thermodynamics to gases to try to understand more about the relationship between energy and temperature through the concept of heat capacity. When a gas is heated, we can use the **molar heat capacity**, the heat capacity per unit mole, to calculate the increase in temperature of the gas. As we shall see, we actually require two different quantities depending on the conditions under which the

© MoE: FDRE Ethiopia Not to be republished

heating takes place: the **molar heat capacity at constant pressure**, C_p, and the **molar heat capacity at constant volume**, C_V. Both have units of J/mol K.

Discussion activity

Why do we only require one value for the molar heat capacity of a liquid or a solid?

Note...

Lower case "c" is commonly used for specific heat capacity, heat capacity per unit mass, whereas upper case "C" is used for molar heat capacity.

When n moles of gas are heated resulting in a temperature rise of ΔT we can write:

$$C_p = \frac{\Delta Q_p}{n\Delta T}$$

and

$$C_V = \frac{\Delta Q_V}{n\Delta T}$$

where ΔQ_p is the energy transferred to the gas at constant pressure by heating and ΔQ_V is the energy transferred to the gas at constant volume by heating. If the gas is kept at constant pressure (approximately true for small changes in volume) then we know that it expands when heated and so does work on its surroundings. If the gas is kept at constant volume then no work is done by the gas. Using the first law of thermodynamics:

$$\Delta U = \Delta Q + \Delta W = \Delta Q - p\Delta V$$

so

at constant pressure: $\Delta Q_p = \Delta U + p\Delta V$

at constant volume: $\Delta Q_V = \Delta U$

From the above equations it can be seen that in order to produce an equal increase in internal energy, and a subsequent increase in temperature, a greater amount of energy must be transferred by heating a gas at constant pressure than by heating a gas at constant volume. This is because, at constant pressure, energy must also be supplied to enable the gas to do work in, for example, pushing a piston back against a restraining force. It is therefore always true that:

$$C_p > C_V$$

Quantitatively, we can also write:

$$\Delta Q_p = \Delta Q_V + p\Delta V$$

and substituting in our definitions of molar heat capacities we obtain:

$$C_p n\Delta T = C_V n\Delta T + p\Delta V$$

© MoE-HDR Ethiopia Not to be republished

We will derive the ideal gas equation in the next section but it is:

$$pV = nRT \quad \text{(where } R = \text{molar gas constant)}$$

If the pressure is constant then a change in temperature of an ideal gas will produce a change in volume given by:

$$p\Delta V = nR\Delta T$$

substituting this we get: $C_p n\Delta T = C_V n\Delta T + nR\Delta T$

which reduces to: $C_p - C_V = R$

This relationship is called **Mayer's equation.**

KEY WORDS

Mayer's equation *the difference between the specific heat of a gas at constant pressure and its specific heat at constant volume is equal to the molar gas constant*

So for gases that behave in an ideal fashion, C_p is always larger than C_V by an amount equal to the molar gas constant, 8.3 J/mol K. Table 1.2 provides the molar heat capacities of some real gases, showing increasing departure from ideal behaviour.

We will show in the next section that the increase in internal energy of one mole of an ideal gas when heated at constant volume, producing a temperature increase of ΔT, is equal to $\frac{3}{2}R\Delta T$. We have already seen that $\Delta U = \Delta Q_V = C_V n\Delta T$ so, for one mole of gas with constant volume, the increase in internal energy is also equal to $C_V\Delta T$. Equating these two expressions we see that for an ideal gas $C_V = \frac{3}{2}R$ and hence $C_p = \frac{5}{2}R$. The ratio of C_p to C_V is commonly written as the symbol γ. For an ideal gas:

$$\gamma = \frac{C_p}{C_V} = \frac{5}{3}$$

Worked example 1.2

When heating one mole of a certain gas at constant volume, an energy transfer of 200 J produces an increase in temperature of 10 K. What temperature increase would there be if the same amount of gas was heated in the same way at constant pressure?

1. As the amount of energy transferred by heating is the same we can say that $\Delta Q_V = \Delta Q_p$ so:

   $$C_v n\Delta T_v = C_p n\Delta T_p$$

   where ΔT_v and ΔT_p are the temperature increases of the gas at constant volume and constant pressure, respectively.

2. Rearranging this equation we obtain:

   $$\frac{\Delta T_p}{\Delta T_v} = \frac{C_v}{C_p} = \frac{3}{5}$$

   and so the temperature increase of the gas at constant pressure is 0.6 x 10 K or 6 K.

© MoE-Ethiopia Not to be republished

Gas	c_p/Jmol^{-1} K^{-1}	c_v/Jmol^{-1} K^{-1}	(c_p – c_v)/Jmol^{-1} K^{-1}
Nitrogen	29.0	20.7	8.3
Hydrogen	28.1	19.9	8.2
Carbon dioxide	36.6	28.2	8.4
Chlorine	34.2	25.1	9.1
Ammonia	37.2	27.8	9.4

Table 1.2 *Molar heat capacities of some real gases. Nitrogen shows behaviour that is closest to that of an ideal gas.*

As discussed, for an ideal gas:

$$\gamma = \frac{C_p}{C_v} = \frac{5}{3}$$

This equation is also valid for monatomic gases like helium and neon. However for diatomic gases such as nitrogen (N_2) and oxygen (O_2) this value slightly different.

	Monatomic (ideal) gases	Diatomic gases
$\gamma = \frac{C_p}{C_v}$	5/3 = 1.67	7/5 = 1.4

For a reversible, adiabatic process, it is also true that:

$TV^{\gamma-1}$ = constant

and

pV^{γ} = constant

where

T = absolute temperature in kelvin

V = volume of gas in m^3

p = pressure of gas in Pa

Therefore it follows that:

$T_1V_1^{\gamma-1} = T_2V_2^{\gamma-1}$

and

$p_1V_1^{\gamma} = p_2V_2^{\gamma}$

© MoE, FDR Ethiopia Not to be republished

Worked example 1.3

A piston contains oxygen at 280 K occupying a volume of 0.25 m^3. The cylinder is compressed adiabatically to 0.14 m^3, find the increase in temperature of the gas.

Oxygen is a diatomic gas so $\gamma = 1.4$, using $T_1V_1^{\gamma-1} = T_2V_2^{\gamma-1}$ substituting in known values gives:

- $(280 \times 0.25)^{1.4-1} = (T_2 \times 0.14)^{1.4-1}$
- $\sqrt[(1.4-1)]{70} = \sqrt[(1.4-1)]{0.14T_2}$ *Solving the left hand side and rooting by the power*
- $70 = 0.14T_2$
- $T_2 = \dfrac{70}{0.14}$ *Rearranging to make T_2 the subject gives*
- $T_2 = 500$ K

Summary

- The first law of thermodynamics states that: $\Delta U = \Delta Q + \Delta W$, where ΔU is the *increase* in internal energy of the system, ΔQ is the amount of energy transferred to the system by heating it and ΔW is the amount of work done *on* the system.
- The internal energy of a gas can be increased by heating it or by doing mechanical work on it.
- When heating a system: $\Delta Q = mc\Delta T$
- When doing mechanical work on a gas: $\Delta W = -p\Delta V$
- In an isothermal process the temperature of the system remains constant.
- In an adiabatic process there is no transfer of energy into or out of the system by heating.
- For any gas: $C_p > C_V$.
- For an ideal gas: $C_p - C_v = R$ and $\dfrac{C_p}{C_V} = \dfrac{5}{3}$.

Review questions

1. In points 2 and 3 on page 12 (how to calculate the increase in internal energy of a system), the processes were adiabatic. How could this have been nearly achieved in practice in the two example situations?
2. Some fire extinguishers contain carbon dioxide stored under very high pressure. Use the first law of thermodynamics to explain why some carbon dioxide solidifies as 'dry ice' when released from such a fire extinguisher.
3. A lump of lead with mass 0.50 kg is dropped from a height of 20 m onto a hard surface. It does not rebound but remains there at rest for a long period of time.

   What are:

   a) ΔQ, (b) ΔW and (c) ΔU for the lead during this process?

   d) What is the temperature change in the lead immediately after the impact? [specific heat capacity of lead = 128 kJ/kg K]
4. 1000 cm^3 of air at 20 °C and 101.35 kPa is heated at constant pressure until its volume doubles.

   a) Use the ideal gas equation to calculate the final temperature of the gas.

   b) Calculate the work done by the gas as it expands.

© MOE Not to be republished

1.3 Kinetic theory of gases

By the end of this section you should be able to:

- State the assumptions made to define an ideal gas.
- Describe the kinetic theory of gases, including the importance of Brownian motion and diffusion.
- Define of r.m.s. velocity of a gas and the mean free path for a gas particle.
- Use the expression for the pressure of an ideal gas in terms of its density and mean square speed of molecules to solve problems.
- Solve problems to determine P, V, T or r.m.s. speed of gas molecules for an ideal gas, given relevant data.
- State Graham's law of diffusion and use it to solve related problems.
- State Dalton's law of partial pressure and use it to solve related problems.

As we have started to look at gases in more detail at the end of the previous section, and before we move on to understand the workings and implications of the second law of thermodynamics, we will take an in-depth look at gases. In particular, we will derive equations that describe the macroscopic quantities and then the microscopic quantities of a gas. By quantitatively comparing these two quantities some remarkably simple and satisfying relationships will be found. In order to allow us to do this, we must make various assumptions about the gas, namely that it is 'ideal'. This is nearly true for a large number of gases as long as the pressure is not too high and the temperature is not too low.

What are the gas laws and the ideal gas equation?

From Figure 1.15 overleaf, the three gas laws may be summarised as follows;

Boyle's law: pV = constant (if T is constant)

Charles' law: $\frac{V}{T}$ = constant (if p is constant)

Pressure law: $\frac{p}{T}$ = constant (if V is constant)

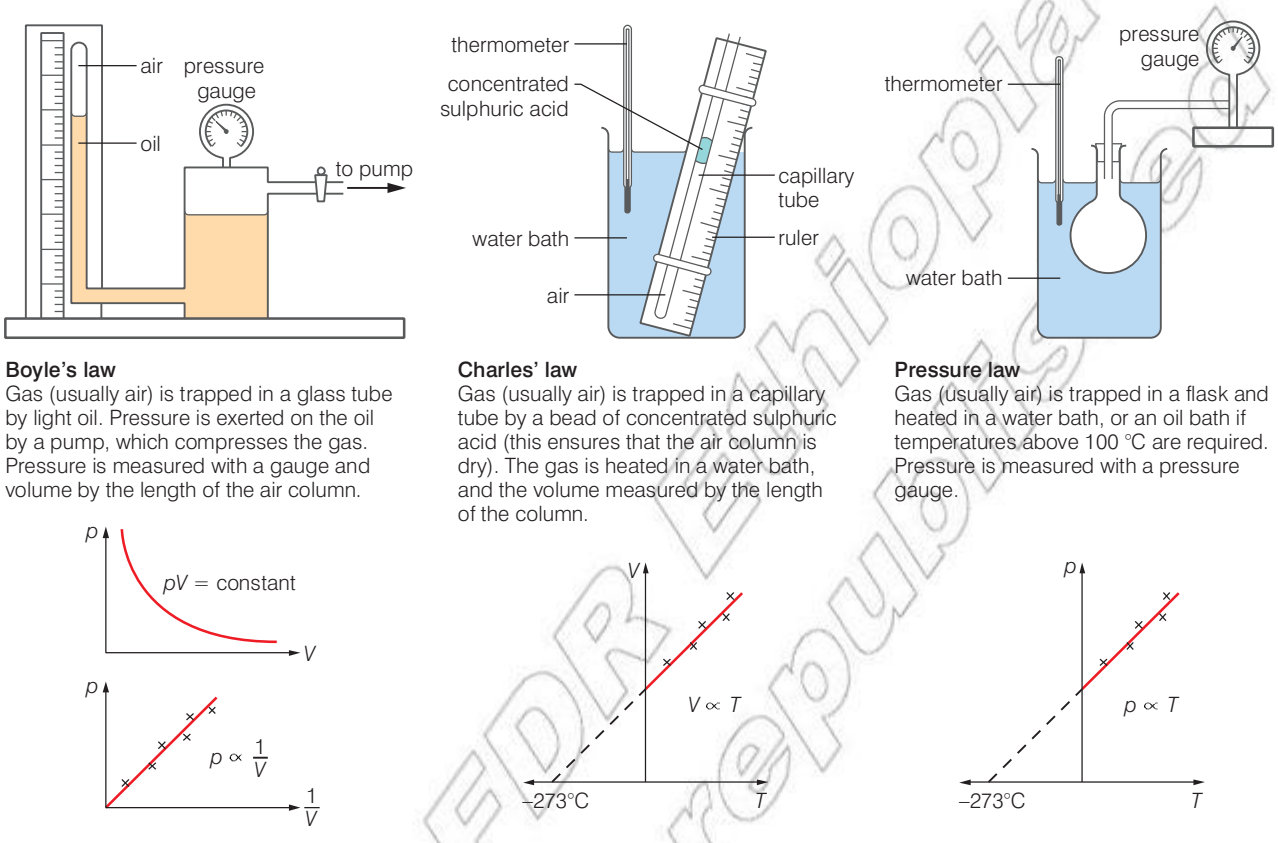
It is worth remembering that:

- these are empirical laws, obtained by experiment, and
- any calculation in thermodynamics must contain temperature given in kelvin, that is, we must use the ideal gas temperature scale at all times.

© MoE-EIO Ethiopia Not to be republished

Combining the three gas laws gives:

$$\frac{pV}{T} = \text{constant}$$



Boyle's law
Gas (usually air) is trapped in a glass tube by light oil. Pressure is exerted on the oil by a pump, which compresses the gas. Pressure is measured with a gauge and volume by the length of the air column.

Charles' law
Gas (usually air) is trapped in a capillary tube by a bead of concentrated sulphuric acid (this ensures that the air column is dry). The gas is heated in a water bath, and the volume measured by the length of the column.

Pressure law
Gas (usually air) is trapped in a flask and heated in a water bath, or an oil bath if temperatures above 100 °C are required. Pressure is measured with a pressure gauge.

Figure 1.15 *Investigating the gas laws*

KEY WORDS

molar gas constant *a constant value given by the pressure times the volume of a gas divided by its absolute temperature*

Avogadro's constant *the number of particles in one mole of a substance*

Experiments also show that for a given pressure and temperature, doubling the amount of gas doubles the volume. The constant may therefore be written as nR where n is the number of moles of particles in the gas and R is the **molar gas constant**, which has a value of approximately 8.3 J/mol K. One mole of 'something' represents 6.0×10^{23} of that something and this number is also referred to as **Avogadro's constant**, N_A.

The equation of state for an ideal gas describes for us the way that a fixed amount of gas behaves as the macroscopic quantities of pressure, volume and/or temperature vary and it is written as:

$$\frac{pV}{T} = nR$$

or

$$\boldsymbol{pV = nRT}$$

© MOE, FDR Ethiopia
Not to be republished

Worked example 1.4

At sea level, atmospheric pressure is 1.0×10^5 Pa, temperature is 300 K and the density of air is 1.2 kg/m^3. What is the density of air at the top of Mount Everest where the temperature is 250 K and atmospheric pressure is 3.3×10^4 Pa?

1. The solution is made easier by dealing with a fixed mass of gas, say 1 kg.

   At sea level: volume of 1 kg of air $= \frac{1 \text{ kg}}{1.2 \text{ kg/m}^3}$

   $= 0.83 \text{ m}^3$

   so: $p_1 = 1.0 \times 10^5$ Pa $\quad T_1 = 300$ K $\quad V_1 = 0.83$ m^3

2. At the top of Everest:
   $p_2 = 3.3 \times 10^4$ Pa $\quad T_2 = 250$ K $\quad V_2$ = to be calculated

3. As the mass of gas stays fixed (we are dealing with 1 kg of air in each case) we can write the ideal gas equation as:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}$$

4. Substituting in all the above values we obtain: $V_2 = 2.1$ m^3. This means that the density of our 1 kg mass of air is 1 kg/ 2.1 m^3 which is 0.48 kg/m^3.

Dalton's law of partial pressures

The equation of state of an ideal gas also explains two laws that were formulated in the early 19th century. Firstly, **Avogadro's law** states that "equal volumes of all gases at the same temperature and pressure contain the same number of molecules".

Secondly, **Dalton's law of partial pressures** states that "the total pressure of a mixture of gases, which do not interact chemically, is equal to the sum of the partial pressures, i.e. to the sum of the pressure that each gas would exert if it alone occupied the volume contacting the mixture." To show this is the case, suppose we have a volume V that contains n_1 moles of a gas with a partial pressure p_1 and n_2 moles of a gas with a partial pressure p_2. If the gas mixture is in thermal equilibrium at temperature T then:

$$p_1 V = n_1 RT \quad \text{and} \quad p_2 V = n_2 RT,$$

and dividing these equations gives $\frac{p_1}{p_2} = \frac{n_1}{n_2}$

By substituting in Dalton's law in the form $p = p_1 + p_2$ we obtain equations for the two partial pressures:

$$p_1 = p\left(\frac{n_1}{n_1 + n_2}\right) \text{ and } p_2 = \left(\frac{n_2}{n_1 + n_2}\right)$$

KEY WORDS

Avogadro's law *equal volumes of all gases at the same temperature and pressure contain the same number of molecules*

Dalton's law of partial pressures *the total pressure of a mixture of gases, which do not interact chemically, is equal to the sum of the pressures that each gas would exert if it alone occupied the volume containing the mixture*

Activity 1.5: Partial pressures

If air is taken to consist of 80% nitrogen and 20% oxygen, what are the partial pressures of these two gases at an atmospheric pressure of 101 kPa?

© Wonderful Ethiopia Not to be republished

KEY WORDS

Brownian motion *the apparently random movement of particles in fluids, caused by the impacts of molecules of the fluid*

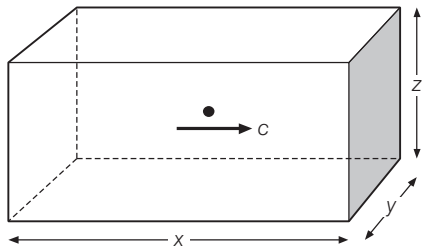


Figure 1.16 *Our starting point for deriving a mathematical relationship between the motion of particles in a gas and the pressure they exert*

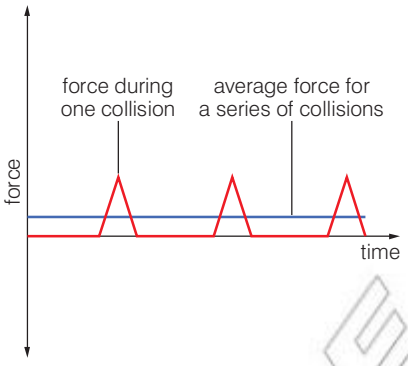


Figure 1.17 *The impulse on the shaded wall is represented by the area under a force–time graph. This is the same whether we consider a series of instantaneous forces or a small but continuous force.*

The kinetic theory of gases

Since 1827, when Robert Brown observed the so-called **Brownian motion** of pollen grains suspended in water, it has been known that liquids and gases contain tiny particles that are moving at high speed. Nowadays, the dancing, zigzag motion of smoke particles as they are bombarded by light and very fast air molecules can be easily observed through a microscope. Einstein showed in 1905 that this motion provided evidence for the existence of atoms; it also leads us to look at gases in a more microscopic and methodical way, to see what we can find out about the kinetics of the tiny particles that make up a gas.

Step 1: To start with, consider a single particle, mass m, travelling at velocity c in a box as shown in Figure 1.16. The particle collides elastically with the shaded wall and rebounds with a velocity $-c$. The change in the particle's momentum is $2mc$.

Step 2: The particle bounces backwards and forwards within the box, exerting a force on the shaded wall many times but only for a short duration each time. As the box has a length X, the time between two successive collisions with the shaded wall is equal to $\frac{2x}{c}$.

Step 3: Newton's second law states that force is equal to the change of momentum divided by the time taken. In each time interval $\frac{2x}{c}$ the wall will cause a change in momentum of the particle of $2mc$. The average force on the particle by the wall is therefore $\frac{mc^2}{x}$.

Step *4*: Newton's third law tells us that as the wall has exerted a force on the particle, the particle has exerted an equal and opposite force on the wall. The average force exerted on the wall over a series of impacts therefore equals $\frac{mc^2}{xyz}$. See Figure 1.17.

Step 5: As pressure is equal to force divided by area and the area of the shaded wall is xz, the average pressure on the shaded wall due to the motion of the single particle is $\frac{mc^2}{V}$. The volume, V, of the box is equal to xyz so the pressure can be written as $\frac{mc^2}{x}$.

Step 6: In reality, there will be N particles in the box moving in random directions. On average, one-third of the particles will be colliding with the shaded wall and the opposite wall and a third will be colliding with each of the other pairs of facing walls. The total average pressure on the shaded wall is therefore $\frac{Nmc^3}{3V}$.

A much more rigorous statistical approach should be used but the result of such an analysis provides the same result as the one that we have used here. As this expression is independent of direction it gives us the total average pressure on any of the walls that make up

© Free State DBE Not to be republished

the container of the gas and it now makes sense to refer to particle speeds rather than velocities.

Step 7: The total mass of the gas is Nm so we can replace $\frac{Nm}{3}$ with the density of the gas, ρ. The total average pressure is now $\frac{\rho c^2}{3}$.

Step 8: We need one final, important alteration. As we have a collection of a very large number of particles there will exist a range of speeds for the particles. As it is the mean pressure that we can measure for any gas then our expression can only provide us with information on the mean of all the squares of the speeds of the particles. We call this the **mean square speed**, $<c^2>$.

KEY WORDS

mean square speed *the average value of the squares of the speeds of particles in a gas*

To conclude, the pressure, p, of an ideal gas can be written in terms of the microscopic quantities of the gas as:

$$p = \tfrac{1}{3}\rho<c^2>$$

where ρ = the density of the gas and $<c^2>$ = the mean square speed of all the particles.

What are the assumptions made in the kinetic theory of gases?

The kinetic theory of gases applies only to ideal gases that have the following properties:

1. The internal energy of the gas is made up of random kinetic energies of particles only. There is no potential energy component. Another way of saying this is that there is no interaction between particles or between particles and the wall except during collisions, all of which are instantaneous.
2. Collisions between particles and between particles and walls are perfectly elastic (there is no loss of kinetic energy during any collision).
3. The volume occupied by the particle is negligible compared to the volume of the gas as a whole.
4. The distribution of velocities of particles is random. That is, the average distribution of velocities over time is the same in all directions. This requires that the number of particles, N, is very large.
5. Newton's laws can be applied to all collisions.

Although no gas is ever perfectly ideal, the results of the above model compare well with that of real gases as long as the gas is well away from the conditions at which it would liquefy.

Activity 1.6: Assumptions in the kinetic theory of gases

Discuss with a partner each assumption in turn. Assess the validity of each assumption for a range of different gases at different temperatures (for example, helium at 4 K, the atmosphere at 280 K and a gas burning in air at 600 K).

© MoE-KICD Ethiopia Not to be republished

KEY WORDS

root-mean-square speed *the square root of the mean of all the squares of the speeds of the particles in a gas*

How fast are molecules in the air moving right now?

Using our model we cannot find out information on the velocities of individual particles or even the range of velocities. We can only work out $<c^2>$.

Taking the density of air at approximately 25 °C as 1.18 kg/m^3 and atmospheric pressure at sea level as 101 kPa:

$$<c^2> = \frac{3p}{\rho}$$

$$= \frac{3 \times 1.01 \times 10^5 \text{ Pa}}{1.18 \text{ kg/m}^3}$$

$$= 2.57 \times 10^5 \text{ m}^2/\text{s}^2$$

By taking the square-root of the mean square speed we do not obtain the mean speed but, instead, the **root-mean-square speed**,

$c_{rms} = \sqrt{<c^2>} = 507$ m/s

It is amazing that we can so easily work out a microscopic property of a gas like the root-mean-square speed of its particles just by knowing two macroscopic quantities of the gas (pressure and density). And the speed is rather large. Try to imagine that, on average, air molecules are striking the skin on your nose at this very moment at over 500 m/s!

Activity 1.7: Mean square speed and root mean square speed

Convince yourself that $<c>$ is different from c_{rms} by calculating the value of both quantities for the following group of velocities: 400 m/s, 450 m/s, 750 m/s, 300 m/s, 500 m/s, 600 m/s. You will find that they are similar but c_{rms} is always greater. Why?

Discussion activity

Refer to Table 1.3. Why does it make sense that c_{rms} is always larger than c_{sound}?

Gas	c_{rms} (m/s)	c_{sound} (m/s)
Helium	1500	1020
Nitrogen	567	356
Carbon dioxide	452	270

Table 1.3 *Values of the root-mean-square speed of particles, c_{rms}, and the speed of sound, c_{sound}, for a number of gases at room temperature and pressure.*

© MoE-Ethiopia Not to be republished

Brownian motion and diffusion

Brownian motion is the apparently random movement of particles in a fluid. It was first observed by the Scottish botanist Robert Brown. He observed pollen grains suspended in water would 'jiggle' around when seen under a microscope.

The first good explanation for this motion was suggested by J. Desaulx in 1877. He said:

"In my way of thinking the phenomenon is a result of thermal molecular motion in the liquid environment (of the particles)."

This turned out to be true. The particle in the fluid (liquid or a gas) was being constantly bombarded from all sides by molecules of the fluid. If the particle is very small, because the impacts are seemingly random on occasion it gets more hits on one side than the other. This gives rise to a net force and causes the particle to accelerate in the direction this net force. These small movements are what make up Brownian motion.

The term **mean free path** (often given the symbol λ) of the particle was used to describe the average distance covered by the particle between successive impacts. In a gas, the lower the pressure of the gas, or the lower the temperature, the greater the mean free path.

KEY WORDS

mean free path *the average distance covered by a particle between impacts in Brownian motion*

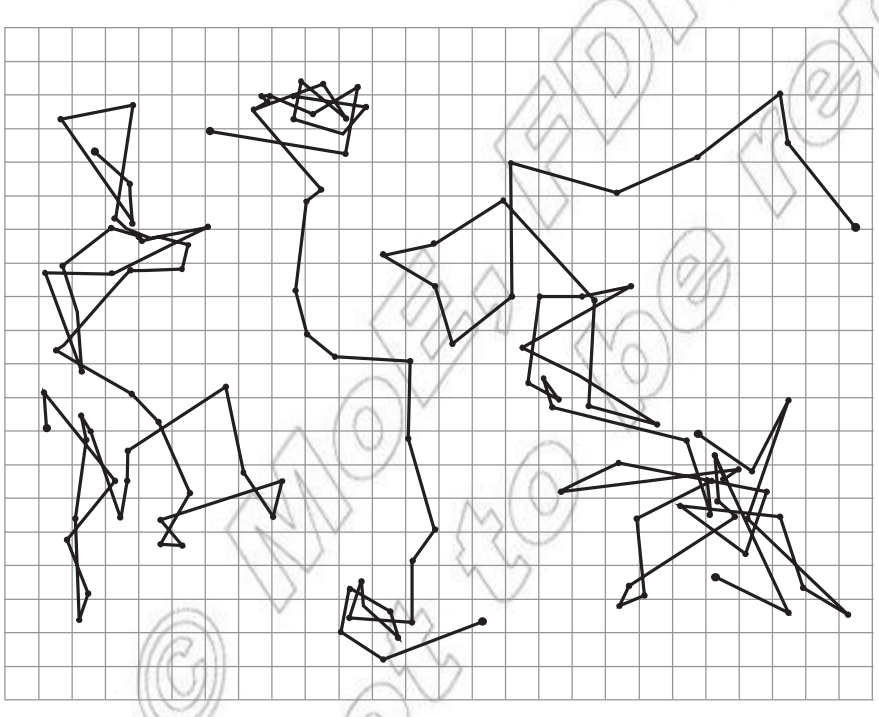


Figure 1.18 *Three tracings of the motion of particles seen under the microscope*

Einstein was the first to come up with an acceptable mathematical model for these movements. He was able to show that this motion can be predicted directly from the kinetic model of thermal equilibrium. This was confirmed experimentally, providing clear evidence for not only the atomic/particle nature of fluids but also the statistical nature of the second law of thermodynamics.

© MoE-EDRE Ethiopia Not to be republished

KEY WORDS

diffusion *the process where particles in a fluid spread from a region of high concentration to a lower one*

Graham's law *the rate of diffusion of a gas is inversely proportional to the square root of its density*

Einstein also produced a statistical analysis of **diffusion**. This is the process where particles in a fluid spread from a region of high concentration to a lower one. You can see this if you add a drop of orange juice to a glass of water. The concentrated orange drop slowly diffuses throughout the rest of the water. Einstein was able to provide a mathematical explanation for this process.

The ideas developed from Einstein's ideas on Brownian motion and diffusion lead to the kinetic theory models already discussed earlier in this chapter.

Graham's law of diffusion

The r.m.s. speed of molecules has an effect on how fast they diffuse through, for example, a cotton wool plug. It is not the velocity of the bulk motion of the gas that is important here but instead the diffusion of the gas through the porous plug is a result of the motion of individual molecules within the gas. **Graham's law** states that "the rate of diffusion of a gas is inversely proportional to the square root of its density", which we will explain below.

Using $pV = nRT$, two gases at a given temperature and pressure with equal volumes of a gas contain equal numbers of molecules so the density of each gas is just directly proportional to the molecular mass of its constituent particles. For two different gases denoted by subscripts 1 and 2: $m_1/m_2 = \rho_1/\rho_2$. From kinetic theory, the molecules in each gas at the same temperature possess the same average kinetic energy, so:

$$\frac{\langle c_1^2 \rangle}{\langle c_2^2 \rangle} = \frac{m_2}{m_1}$$

and therefore:

$$\frac{\langle c_1^2 \rangle}{\langle c_2^2 \rangle} = \frac{p_2}{p_1}$$

leading to:

$$\frac{\sqrt{\langle c_1^2 \rangle}}{\sqrt{\langle c_2^2 \rangle}} = \sqrt{\frac{p_2}{p_1}}$$

This explains why the rate of diffusion in a gas – which is dependent on the r.m.s. speed of molecules – is inversely proportional to the square root of the density of the gas.

© MoE. Not to be republished

Worked example 1.5

The r.m.s. speed of molecules in a container of helium gas at 332 K is 1440 m/s. What is the r.m.s. speed of molecules of hydrogen gas in the same container at he same temperature and pressure?

Avogadro's law tells us that the number of molecules of hydrogen is the same as the number of molecules of helium. The ratio of densities of the two gases therefore depends only on the ratio of the molecular masses. The mass of a helium molecule is twice the mass of a hydrogen molecule so $p_H = p_{He}/2$. Therefore:

r.m.s. speed of molecules in the hydrogen gas = $\sqrt{2} \times 1440$ m/s

= 2036 m/s

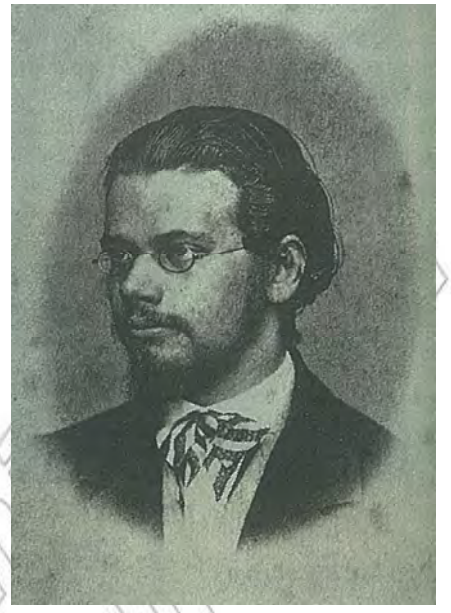


Figure 1.19 *Ludwig Boltzmann (1844–1906)*

What if we compare these two world views?

We now have two equations that describe the behaviour of a gas. The first is a result of experimental study of gases. The second is based on a model of the **kinematics of particles** in the gas. Let us see what else we can find out about gases from these two equations:

Equation 1: $pV = nRT$

Equation 2: $p = \frac{1}{3}\rho\langle c^2\rangle = \frac{\frac{1}{3}Nm\langle c^2\rangle}{V}$

Equating pV in both Equation 1 and Equation 2:

$$\frac{1}{3}Nm\langle c^2\rangle = nRT$$

From the definition of a mole: $N = nN_A$, where $N_A = 6.0 \times 10^{23}$, so:

$$\frac{1}{3}m\langle c^2\rangle = \frac{RT}{N_A}$$

Defining the average kinetic energy of a particle as $\langle E_k\rangle = \frac{1}{2}m\langle c^2\rangle$, we can state that:

$$\langle E_k\rangle = {}^3/_2 kT \text{ (where } k = \frac{R}{N_A}\text{)}$$

The constant k is called the **Boltzmann constant** and it is effectively the gas constant for one particle in a gas (as opposed to R which is the gas constant for one mole of gas). The value of k is 1.38×10^{-23} J/K.

Again, we have an expression that can tell us a microscopic quantity of a gas, the average kinetic energy of its particles, from, in this case, a single macroscopic quantity, the Kelvin temperature. This seems remarkable until you go back and see how we originally defined temperature in section 1.1. However, we must remember that the above quantitative relationship is only true for an ideal gas.

KEY WORDS

kinematics of particles *study of the motion of particles within a gas*

Boltzmann constant *the gas constant for one particle in a gas*

© MOEEDR Free to be reproduced

DID YOU KNOW?

The Boltzmann constant is named after an Austrian physicist named Ludwig Boltzmann who contributed a great deal to the study of thermodynamics, including the work on the kinetic theory of gases. Unfortunately, Boltzmann's theories of the behaviour of matter were not generally accepted at the time, even ridiculed, and he committed suicide. Seventy years after his death, Zartmann and Ko showed that the distribution of velocities in the molecules of a gas agreed with his predictions, as shown in Figure 1.20.

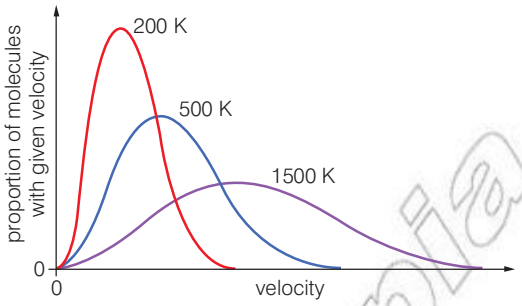


Figure 1.20 *The distribution of molecular speeds within a gas changes and the average speed of molecules increases as the temperature of the gas increases.*

In addition, as this is an ideal gas the internal energy is equal to the sum of all the random kinetic energies of its particles. The internal energy of one mole of gas, U, can therefore be written as:

$$U = N_A\langle E_k\rangle = {}^3/_2 RT$$

Summary

- The ideal gas equation: $pV = nRT$.
- With all the assumptions of an ideal gas, the kinetic theory of gases states that $p = {}^1/_3 \rho \langle c^2\rangle$.
- The r.m.s. speed is the square root of the mean of all the squares of the speeds.
- The average kinetic energy of a molecule in an ideal gas $= {}^3/_2 kT$.

Review questions

1. An ideal gas is sealed within a container at a temperature of 17 °C and a pressure of 101 kPa. The container is heated until the temperature of the gas reaches 100 °C. A valve in the container is then opened to allow gas to escape until the pressure falls back to 101 kPa at 100 °C.
   a) Calculate the pressure in the gas just before the valve is opened.
   b) Calculate the fraction of the initial mass of gas that was lost as a result of opening the valve.
2. Calculate:
   a) the average kinetic energy of a nitrogen molecule at 20 °C
   b) the average kinetic energy of an oxygen molecule at 20 °C
   c) the internal energy of one mole of nitrogen gas at a pressure of 50 kPa occupying a volume of 5.0×10^4 cm^3.
3. Estimate the number of impacts that air molecules make with the palm of your hand each second, assuming that air consists entirely of nitrogen (molar mass = 0.028 kg).

© MoE Ethiopia Not to be republished

1.4 Second law of thermodynamics, efficiency and entropy

By the end of this section you should be able to:

- State the second law of thermodynamics.
- Appreciate that the second law of thermodynamics places sharp constraints on the maximum possible efficiency of heat engines and refrigerators.
- Distinguish between reversible and irreversible processes.
- Define entropy as a measure of disorder and state the second law of thermodynamics in terms of entropy.

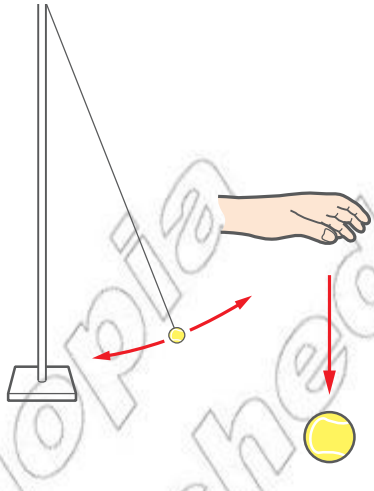


Figure 1.21 *Considering the direction of time: (a) a simple pendulum in motion; (b) a tennis ball falling towards the floor and bouncing a number of times*

Imagine we viewed a short film of the two situations shown in Figure 1.21. Now imagine we viewed the same two films running backwards. Would we notice any change?

For the pendulum, we would not notice any difference. Energy would still flow from the gravitational potential form to the kinetic form and back again. It would appear that the physics of the pendulum system is the same whether time moves forwards or backwards.

However, watching a tennis ball bouncing to ever greater heights would seem very odd indeed. Despite satisfying the law of conservation of energy, the idea that with every bounce the ball absorbs energy through heating and so rebounds with a greater kinetic energy than it struck the floor with, goes against our everyday experience. Of course, if the film of the pendulum lasted longer it, too, would only make sense with time flowing forwards (the amplitude of the swings cannot increase over time).

This apparent gap in our description of the universe will now be filled by the second law of thermodynamics. The second law provides a necessary direction to time and, importantly, places limits on the efficiency of systems and how they behave.

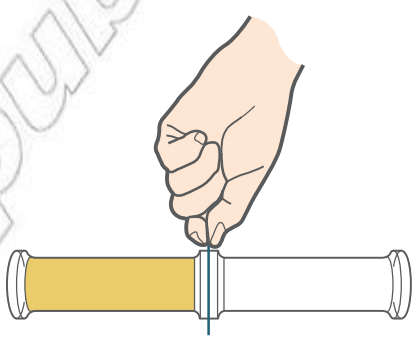


Figure 1.22 *Two glass cylinders separated by a glass cover slide. In the left-hand cylinder is bromine gas. In the right-hand cylinder is air.*

Entropy and the second law of thermodynamics

As we have already seen, the concept of energy cannot be used to determine the possible direction of time as the sum of all energies in an isolated system does not change with time. Instead we need the concept of **entropy**. We will introduce this new quantity using the example of bromine gas in a cylinder as in Figure 1.22.

Imagine that the cover slide is removed and the bromine gas is left to diffuse for a long time. When we next look we expect to see that the bromine gas has spread evenly within both cylinders. We would certainly not expect to see this process happen in reverse, for all the bromine molecules to end up in the left cylinder only. But would it be possible? In order to investigate this possibility let us deal with a simpler situation where there are only five molecules, as shown overleaf in Figure 1.23.

KEY WORDS

entropy *a measure of the amount of disorder in a system*

© MOE-EBK Ethiopia Not to be republished

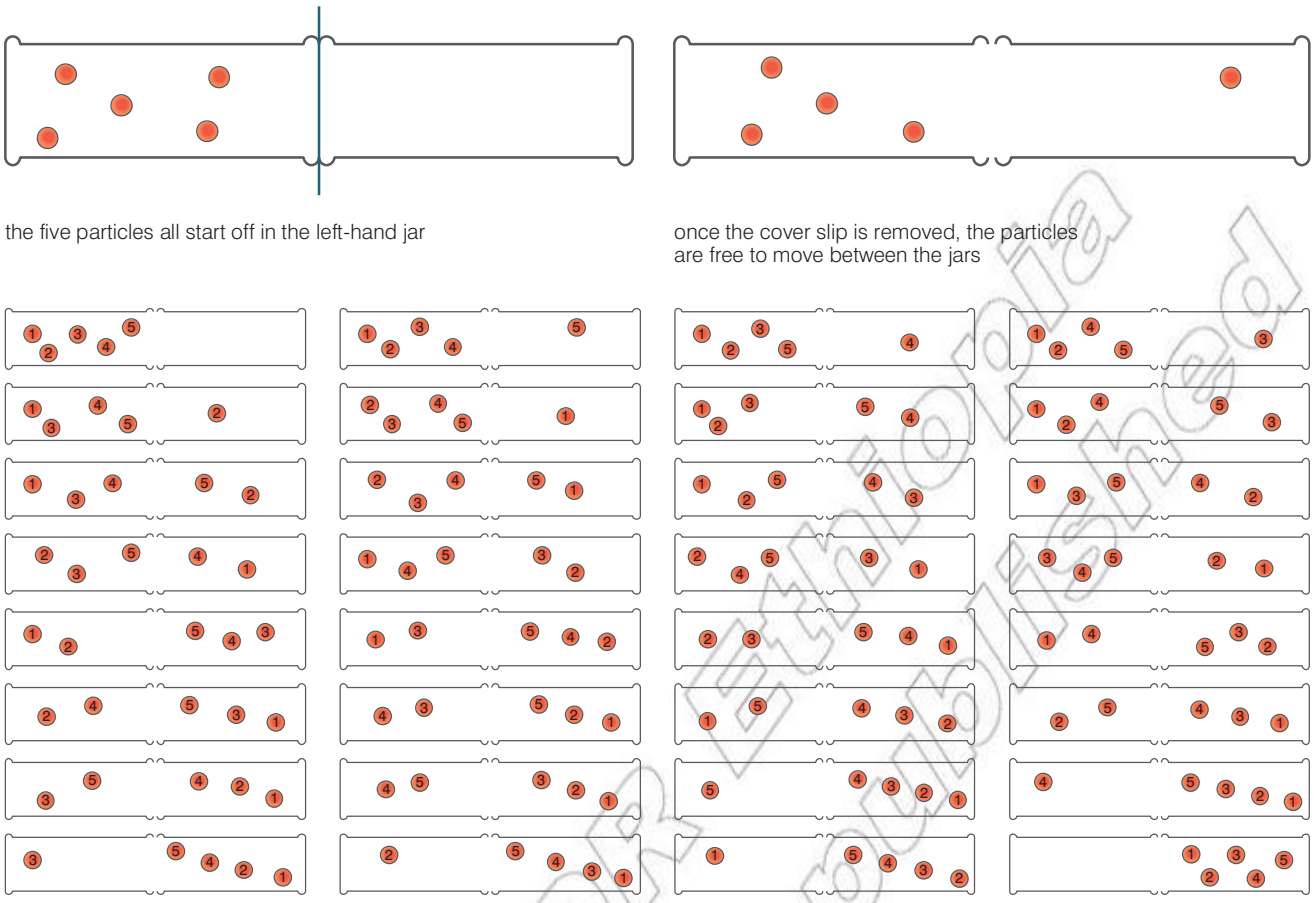


Figure 1.23 *To start with, all five bromine molecules are certain to be in the left-hand cylinder. Once the cover slip is removed each particle is in one of two states: either in the left-hand cylinder or in the right-hand cylinder. The number of possible ways in which the five particles might be distributed between the cylinders is therefore 2 × 2 × 2 × 2 × 2, written as 2^5. The probability of all the particles being in the left-hand cylinder is therefore 1 in 2^5, or 1 in 32.*

KEY WORDS

disorder *chaos or a lack of order or arrangement*

quanta *the set amounts by which atoms can increase or decrease their energy*

Entropy is a measure of the amount of **disorder** (or chaos) in a system and the measure of disorder is equal to the number of ways that the system *may* be arranged. Before the cover slide was removed from our simple system it was very *ordered* as there was only one possible arrangement for the five particles: they were all in the left-hand cylinder. The entropy of the system was small. After the cover slide was removed, the amount of disorder in the system, and hence its entropy, increased as there were then 32 possible arrangements for the particles. The particles may all briefly move back into the left-hand cylinder but the entropy will still be higher as there are still 32 possible arrangements for the particles.

The second law of thermodynamics states that:

"No process is possible in which there is an overall decrease in the entropy of the universe."

In a more realistic situation like our sample of bromine gas in Figure 1.22, there may be of the order of 10^{22} molecules present. This means that after the cover slide is removed, the increase in entropy will be very great. How likely is it that all the bromine molecules will end up back in the left-hand cylinder? The probability is 1 in $2^{10^{22}}$. To put this in context, if we observed electronically the cylinders at a rate of 1 MHz then, on average, in order to observe all the molecules in the left-hand cylinder we would have to wait for a time equivalent to N times the age of the universe. In order to write the number N you would have to write down a "1" and then write down the following group of zeroes a

© Mindset Learn. Not to be republished

billion, billion, billion times: 000,000,000,000,000,000,000,000,000,000. If you think about it, you really would not want to try this experiment for real! So we are justified in saying that the particles in the gases will *always* spread out and that the entropy (and disorder) of the system will *always* increase.

Discussion activity

If W is equal to the amount of disorder in a system, that is, the number of possible arrangements within the system, then the entropy, S, of that system is given by the expression $k \ln(W)$, where k is the Boltsmann constant. In units of J/K, can you work out the entropy of the bromine gas in Figure 1.23 after the cover slide is removed?

How is entropy related to energy and heat?

The same ideas about entropy and the second law of thermodynamics are not just true in prescribing the direction of time when observing particles in a gas. They also describe how photosynthesis occurs, how many machines work (such as photocopiers, for example) and why energy tends to spread out through the process of heating.

Just as the spreading out of particles in Figure 1.23 represented an increase of entropy, the spreading out of random kinetic energy through heating also represents an overall increase in entropy. In the early 20th century it was found that atoms can only increase or decrease their energy in set amounts, called **quanta**. It is the behaviour of quanta, tiny packets of energy, that is similar to the behaviour of particles in a gas. If all the quanta that exist in a system are held by a small number of atoms then the number of possible arrangements of those quanta is small. However, if the same number of quanta were distributed amongst a larger number of atoms then, although the total energy of the system is the same, the number of possible arrangements (and therefore the entropy) increases. This is why energy in an isolated system tends to spread out over time: the entropy of the system must not decrease. This explains why the tennis ball cannot bounce to ever greater heights and why a smashed bowl does not spontaneously turn into an undamaged bowl.

You may suggest that, in Figure 1.23, we somehow pick up the five theoretical particles, one by one, and place them back in the left-hand cylinder and seal it, and that this will reduce the disorder in the cylinders. It will, but the cylinders are not then an isolated system. By doing mechanical work on the particles you must, in the process, have caused a heating effect elsewhere that would have caused a larger increase of entropy than the decrease in entropy of the particles. Ultimately, the system in question encompasses the entire universe, hence the definition of the second law of thermodynamics.

DID YOU KNOW?

Computer processor chips run hot. This is not merely due to resistance heating as a result of the flow of electric current. In addition, heating must occur in order to counteract the decrease in entropy that the chip is causing by processing data: in other words ordering that data.

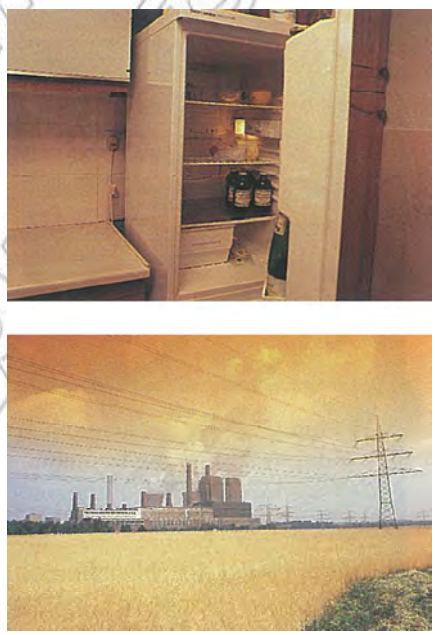


Figure 1.24 *The second law of thermodynamics can also be written as "the spontaneous transfer of energy from a cooler body to a hotter body is not possible". As we shall see in the following section, a refrigerator is an example of a heat pump where, by doing work on the system, energy is transferred from a colder body to a hotter body. This appears to directly contradict the second law. But it does not, as the transfer is not spontaneous. Somewhere else in the world a lump of coal is burned to provide the necessary electricity and the burning coal increases the entropy of the universe by a greater amount than the refrigerator decreases it.*

© Not to be republished

KEY WORDS

heat engine *a device designed to do useful mechanical work*

heat source *a device or body that supplies heat*

heat sink *a body which absorbs heat from other bodies with which it is in thermal contact*

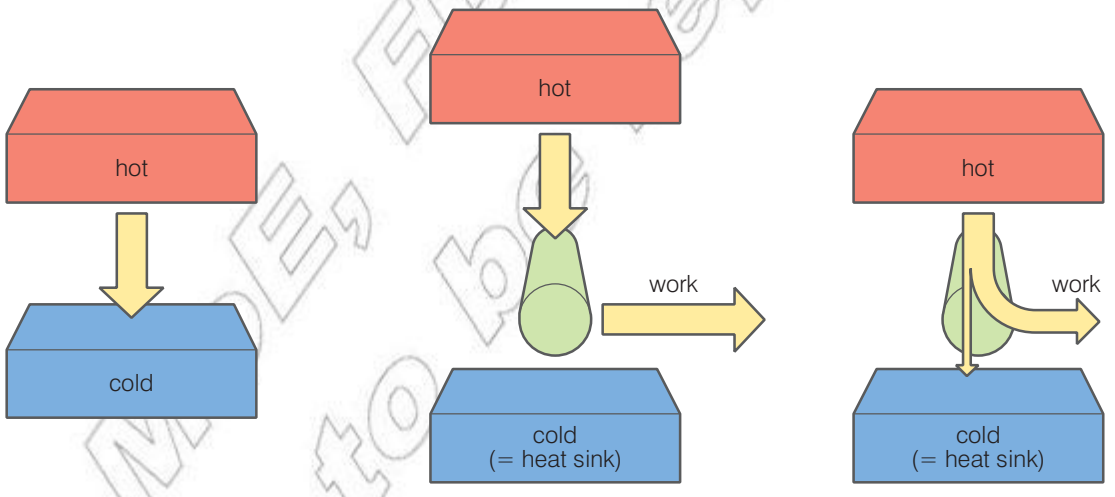
Why can we not build a power station that is perfectly efficient?

As we shall discover in more detail in the following section, even if we removed all sources of friction and electrical resistance from within a power station, its efficiency would still be far less than 100%. This is explained by the second law of thermodynamics.

A **heat engine** is any device that is designed to do useful mechanical work by taking energy from a hot body, often called the **heat source**. A power station is an example of a heat engine. As described in detail in Figure 1.25, the second law of thermodynamics can also be written in the form:

"the complete conversion of energy from a hot source into work is not possible."

Instead, it is necessary to transfer some of the random kinetic energy of the particles to a colder body, often called a **heat sink**. The heat sink is also necessary in order to draw the energy from the hot source. This places very strict limits on the efficiency of heat engines of any kind in our universe. Nothing can be 100% efficient and many heat engines struggle to achieve a theoretical maximum efficiency of even 50%. As we shall soon see, it is the relative temperatures of the heat source and the heat sink that determine how much work can be done as a proportion of the energy input.



When a hot object comes into contact with a cooler one, there is a net flow of energy from the the hot body to the cooler one simply because there are more ways of arranging the quanta when this happens

If energy is removed from a heat source (i.e. a hot body). the entropy of the body decreases. If all this energy is then converted into work, the overall change in entropy is a decrease, which contravenes the second law

If a small amount of energy is allowed to flow into the heat sink, the rise in entropy of the heat sink can offset the drop in entropy of the heat source – so the second law is satisfied

Figure 1.25 *An explanation of why heat engines can never be perfectly efficient. The green cylinder represents the heat engine.*

© MDE, Edrington Not to be reproduced

Discussion activity

The temperature differences that exist between astronomical objects in the universe can be thought of as doing work. For example, the temperature difference between the Sun and the Earth can be considered to power life itself. As the entropy of the universe must always increase what do you think the universe will eventually be like?

What processes can we reverse and why is this important?

When we discussed ideal gases in the previous section, it was clear that no gas is truly ideal but many gases were close enough to ideal, and the assumptions of an ideal gas simplified the theory enough, that the derivation of the kinetic theory of gases was a necessary and worthwhile process. The discussion of thermodynamic reversibility is of similar merit.

- A reversible process is one that can be reversed by means of an infinitesimally small change in a property of the system without a transfer of energy.

Due to the need for an infinitely small change, the system must be at rest during the entire process and the process will therefore take an infinite amount of time. Perfectly **reversible** processes are impossible. However, if the system can respond to a change in its state faster than the change is occurring then a process may be approximately reversible.

For a process to be reversible it must also be right on the limit of the second law of thermodynamics: there must be no change in entropy. For if a process causes an increase in entropy of the system then clearly the exact reverse process could not happen. Despite being a practical impossibility, a reversible heat engine therefore has the maximum possible efficiency allowed by the Second Law of Thermodynamics. It transfers *just* enough energy to the heat sink to maintain the same total entropy and so transfers the maximum amount of useful work.

Just like an angry word said in haste, every process that occurs in the universe is actually **irreversible**, either because it has caused entropy to increase or because a finite change in the system occurred. Milk stirred into coffee does not then spontaneously collect together in the cup. All of the water on a towel does not spontaneously return to the surface of your hand.

When we look at real heat engines we require them to run in a **cycle**. This is where a system is taken through a series of different states and then finally returns to its original state. In this way, the engine can do a large amount of useful work over a large number of repeated cycles. In order to be able to predict the outcome of using

KEY WORDS

irreversible *a process that cannot be reversed, either because entropy has increased or because there has been a finite change in the system*

reversible *a process that can be reversed by means of an infinitesimally small change in a property of the system without a transfer of energy*

Cycle *an interval of time over which a system is taken through a series of different states before returning to its original state*

Activity 1.8: Irreversible processes

With your partner, describe the following processes at a molecular level and hence why they are irreversible: (a) ice cubes melting in a glass of water; (b) a heater warming the air in a room; and (c) clothes drying in a breeze. Remember that if a system did undergo a reversible process, as no change in entropy occurs, then it would be equivalent to time standing still.

© Wits Education Enterprise NOT to be reproduced

a cyclical heat engine, the whole process needs to be reversible, or, at least, nearly reversible. If it was not then the pressure, volume and temperature of the system will be different after each complete cycle and this would make the analysis of the situation very complicated.

Summary

- Entropy is a measure of the amount of disorder in a system.
- The second law of thermodynamics states that no process is possible in which there is an overall decrease in the entropy of the universe.
- A reversible process can be reversed by means of an infinitesimally small change in a property of the system without a transfer of energy.
- As well as defining the direction of time, the second law of thermodynamics implies that no process or cycle is perfectly reversible and no heat engine or refrigerator can be perfectly efficient.

Review questions

1. The concept of 'negentropy' has been suggested – it is a measure of the order in a system. Do you agree with the statement "In our lives, we do not so much consume energy as consume negentropy". Give your reasons.
2. A table tennis ball is placed on the hot plate of an oven hob and gently heated. At some point, all the kinetic energy transferred to the particles in the ball results in all the particles moving upwards at the same time and the ball jumps into the air.
   a) Does this contradict the first law of thermodynamics? Explain why.
   b) Does this contradict the second law of thermodynamics? Explain why.
   c) Is the ball likely to jump in the air? Explain why.
3. When heating occurs at a constant temperature, the change in entropy that occurs is equal to Q/T, where Q is the amount of energy transferred and T is the temperature. Referring to Figure 1.25, can you explain how it is possible for a heat engine to transform some of the energy from the hot source into work and for the total entropy of the system to still increase?

© MoE, Not to be republished

1.5 Heat engines and refrigerators

By the end of this section you should be able to:

- Describe the fundamental principles of heat engines and refrigerators.
- Solve problems involving heat, work and efficiency in a heat engine.
- Identify that all real heat engines lose some heat to their surroundings.
- Investigate the physical principles that all heat engines and refrigerators must obey.

What is a heat engine?

We briefly mentioned heat engines in the previous section. A heat engine is a device that transforms heat energy into mechanical work. Obvious examples include petrol, diesel and jet engines; however, a power plant (such as a nuclear power plant) could be considered to be a heat engine, as could the toy drinking bird in Figure 1.26.



Figure 1.26 *The drinking bird is an example of a heat engine; the temperature difference between the room and the cooler water drives the bird.*

The drinking bird is a heat engine that uses the temperature difference between the room and the water to convert heat energy to mechanical work.

Like all heat engines, the drinking bird works through a thermodynamic cycle (more on these in a moment). The bird starts off vertical with a wet foam head. It then runs through the following process:

1. The water evaporates from the foam head; this evaporation lowers the temperature of the glass head.
2. This temperature drop causes some of the vapour in the head to condense. The lower temperature and condensation cause the pressure to drop in the head (from the ideal gas law).
3. As there is now a small pressure difference between the head and the base, liquid is drawn up the tube.
4. As liquid flows into the head, the bird becomes unbalanced and tips over.
5. When it tips, the head re-enters the water and becomes wet again. At the same time the bottom end of the neck tube rises above the surface of the liquid. A small bubble of vapour rises up the tube through this gap, displacing liquid as it goes. This causes liquid to flow back to the bottom bulb. This equalises the pressure.
6. With the liquid back in the bottom bulb the bird tips back to the vertical; the process then repeats.

© MoE FDRE Ethiopia Not to be reproduced



Figure 1.27 *A more obvious heat engine*

All heat engines require this temperature gradient to function. Often through more complex thermodynamic processes the heat engine extracts some of the thermal energy flowing from a hotter region to a colder one and coverts this into mechanical work. This may be seen in the simple schematic below:

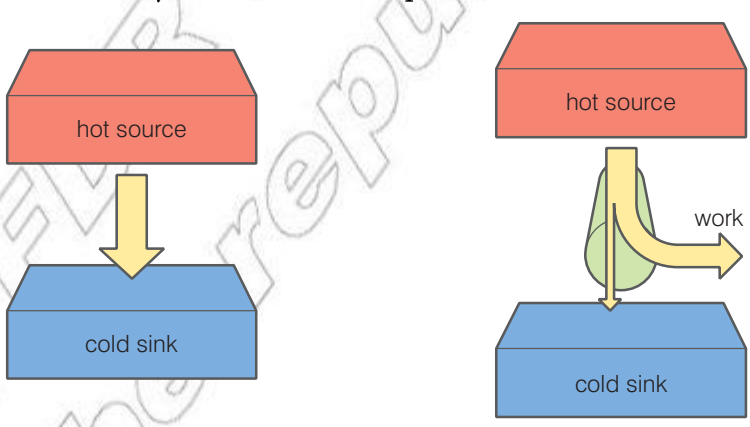


Figure 1.28 *A simple schematic of a heat engine*

Here the heat engine is represented by the green cylinder. Thermal energy is transferred through the engine from a hot source to a cold sink (also referred to as a heat sink). As mentioned previously, the cold sink is necessary in order to 'draw' the energy from the hot source. Processes inside the engine then convert some of this thermal energy into work.

The operation of heat engines is bound within the first and second law of thermodynamics. The first law involves the application of conservation of energy to the system. This often involves work done on or by gases within the engine.

As previously mentioned the second law can also be written in the form:

- The complete conversion of energy from a hot source into work is not possible.

The second law therefore sets limits on the maximum theoretical efficiency of any heat engine (more on this later).

© MoE; Free Ethiopian Not to be republished

How do heat engines work?

Most heat engines use thermodynamic processes and involve gases doing work. Previously in section 1.2 we looked at isobaric, adiabatic and isothermal expansion along with isochoric pressure increases. These processes are important in the operation of most heat engines and may be seen represented on a pressure vs. volume graph, as shown in Figure 1.29.

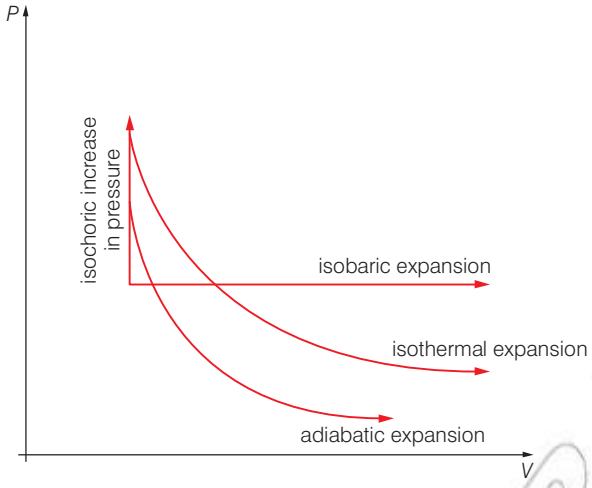


Figure 1.29 *p–V diagrams for various processes*

Table 1.4 *Summary of some thermodynamic processes*

Process	Meaning
Adiabatic	No heat transfer (heating or cooling results from pressure change – work is either done on or by the gas)
Isochoric	Constant volume (also called isometric)
Isothermal	Constant temperature
Isobaric	Constant pressure

Two or more of these processes are combined to form a simple cycle. As this cycle repeats, work is continuously extracted from the heat transferred from the hot source to the cold sink.

A very simple cycle may involve just four stages. This may be seen in Figure 1.30.

If we consider each part of the cycle in turn we can see how work might be extracted from the system.

1 to 2

The volume of the gas inside the engine is reduced at constant pressure (isobaric). This may achieved through a slow compression, allowing heat to flow out of the system or more commonly by rapidly cooling the system.

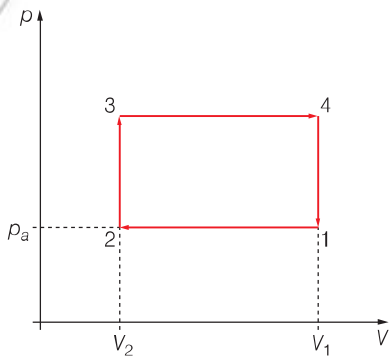


Figure 1.30 *A simplified p–V cycle*

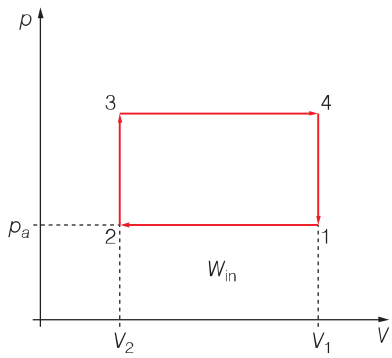


Figure 1.31 *The work done on the gas (W_{in}) as it is compressed at constant pressure*

© ERDR Ethiopia Not to be republished

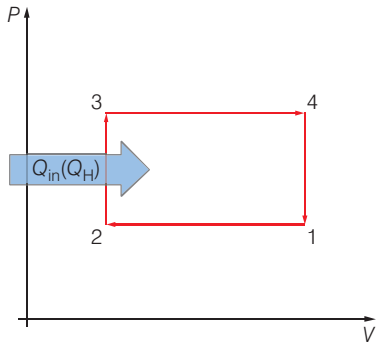


Figure 1.32 *Adding heat causes an isochoric pressure increase.*

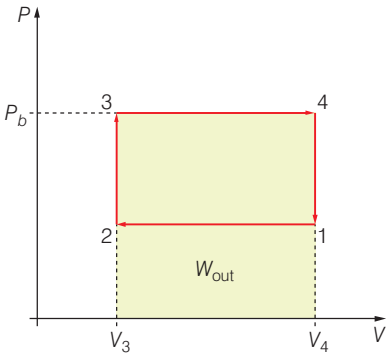


Figure 1.33 *The work done by the gas (W_{out}) as it expands at constant pressure*

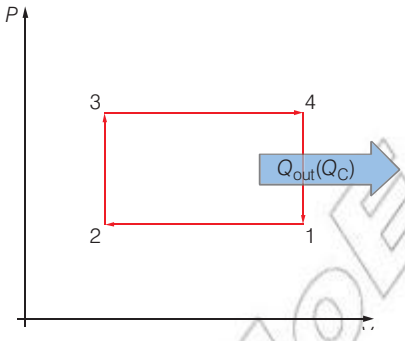


Figure 1.34 *Removing heat causes an isochoric pressure decrease.*

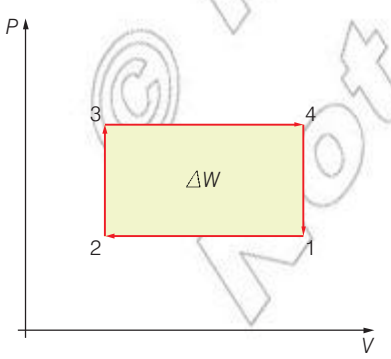


Figure 1.35 *The net work done is equal to the area of the enclosed cycle*

You should recall the work done on the gas at constant pressure is given by:

- $\Delta W = p\Delta V$

In terms of the quantities in Figure 1.31:

- $W_{in} = p_a(V_1 - V_2)$

This is equal to the area under the line 1–2, as shown in the diagram.

As this is an example of an isobaric process there must be no increase in temperature. If we apply the first law of thermodynamics we get:

- $\Delta U = \Delta Q + \Delta W$
- $0 = \Delta Q + p_a(V_1 - V_2)$
- $\Delta Q = -p_a(V_1 - V_2)$

ΔQ is the heat flow out of the system.

2 to 3

Heat is then allowed into the system (Q_H or Q_{in}). This is often achieved by igniting a fuel air mix. This may cause a rapid increase in pressure, whilst the volume remains constant.

Again considering the first law of thermodynamics:

- $\Delta U = \Delta Q + \Delta W$

In this case as there is no change in volume then no work is done on or by the gas and so all of the heat energy goes into increasing the internal energy of the gas:

- $\Delta U = Q_{in}$

This gives rise to an increase in temperature and so an increase in pressure.

3 to 4

The gas may then expand and as it does so it does work on its surroundings. This is very much an idealised example as heat would also need to be put in at this stage in order to keep the pressure constant.

The work done by the gas as it expands is equal to the area under the line 3–4, as shown in Figure 1.33.

4 to 1

Heat energy is then extracted from the system at constant volume. This leads to a drop in pressure and the system returns to its starting point. This process then repeats or 'cycles'. In each cycle, the work extracted is given by the area enclosed by the cycle.

The larger this area the greater the amount of work extracted per cycle.

Remember, this is a very simplified example of how processes may be combined to form a simple cycle.

© MoE FDRE Ethiopia Not to be republished

Examples of real heat engines

Real heat engines combine the different thermodynamic processes in much more complex ways. We shall look at two examples of real engines in more detail: petrol engines and diesel engines. In both cases we will take an idealised engine and discuss the processes involved and how work is extracted from the system.

Petrol engine

Inside a petrol engine fuel is burnt inside a piston. This piston moves up and down and extracts some the thermal energy as mechanical work. This is then usually used to turn a driveshaft and subsequently turn the wheels on a vehicle.

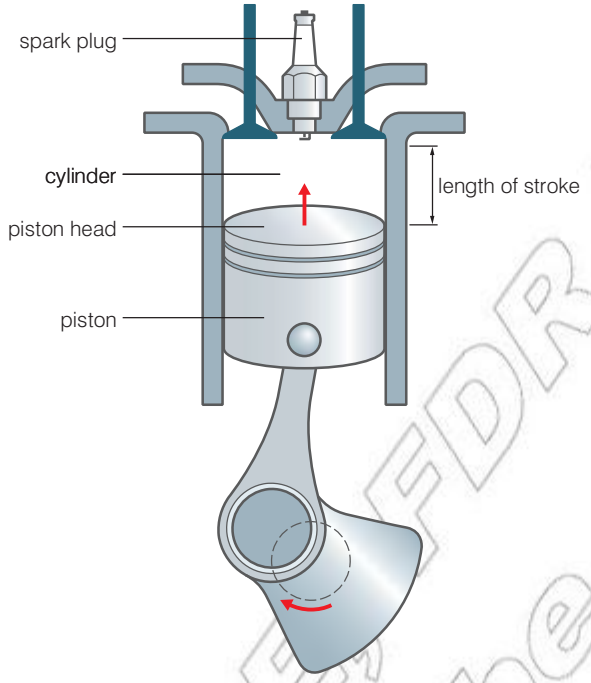


Figure 1.36 *A diagram of a simple piston from a petrol engine*

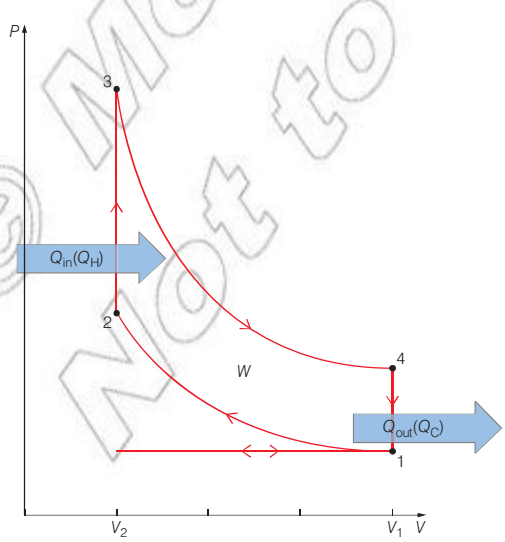


Figure 1.37 *An idealised Otto cycle (petrol engine)*

DID YOU KNOW?

Nikolaus August Otto was a German engineer and inventor. He is credited as the designer of the first internal combustion engine, which efficiently burnt fuel directly inside a piston.

© MoE-FDR Ethiopia
Not to be republished

DID YOU KNOW?

Most petrol engines employ a four stroke process. The strokes are induction, compression, power and exhaust. Can you match them to the ideal Otto cycle p–V diagram?

The cycle responsible for the extraction of work from a petrol engine is referred as the Otto cycle. The diagram in Figure 1.37 shows an idealised p–V diagram for the Otto cycle.

Starting on the bottom left of the diagram the volume of the gas inside the piston is V_2. The piston is pulled down (due to the rotation of the shaft) and a fuel air mix is drawn in through the open inlet valve. The pressure remains constant and the volume of gas inside the cylinder increases to V_1.

The valve is closed and the piston moves back up, compressing the gas back to V_2. This process happens quickly and so the compression is adiabatic, no heat flows out of the gas. This results in an increase in temperature (and so pressure) as work is done on the gas. This can be seen in the line 1–2.

At 2 the fuel air mix is ignited by a spark from the spark plug. Combustion occurs very quickly and as a result the process can be considered to be isochoric, the volume of the gas remains constant. There is a significant amount of heat realised in this process (Q_{in} or Q_H). This heat increases the temperature and so the pressure of the gas inside the piston. This can be seen in the line 2–3.

Between 3 and 4 work is done by the gas as it forces this piston down. This again happens very quickly without any heat flowing out of the system and so it is an adiabatic change. The volume of the gas increases back to V_1 and the pressure drops.

Heat then flows out of the system (Q_{out} or Q_c) and the temperature of the gas drops causing a drop in pressure back to the original pressure. This can be seen in the line 4–1.

The final part of the cycle involves the piston moving back up, but this time the exhaust valve is open and so the pressure remains constant as the waste gases are expelled. We are now back at the bottom left and the process starts again.

Just like our simplified earlier example the difference between the work done by the gas and the work done on the gas is the area enclosed by the cycle. The power of the engine is then found as the product of this difference and the number of cycles per second.

This is very much an idealised Otto cycle. There are no friction losses, the combustion is isochoric (i.e. happens almost instantaneously) and there is no heat entering or leaving the gas between 1–2 and 3–4. In reality there are a number of losses through each part of the cycle. All real heat engines transfer additional thermal energy to their surroundings at every stage.

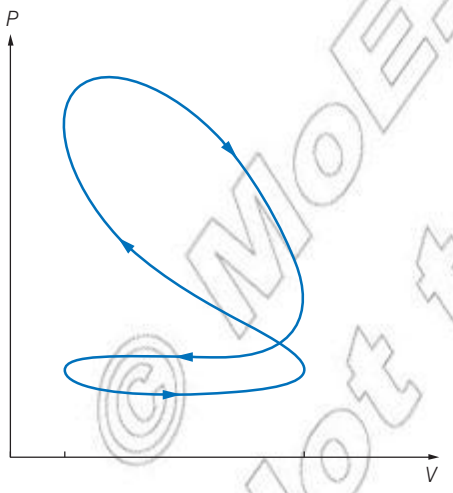


Figure 1.38 *A more realistic Otto cycle*

A more realistic cycle may be seen in Figure 1.38.

Diesel engine

Just like a petrol engine fuel is burnt inside a piston. This piston moves up and down and extracts some the thermal energy as mechanical work. This is then usually used to turn a driveshaft and subsequently turn the wheels on a vehicle. However, there are a few key differences in the process.

© MoE - FDR Electronic Not to be republished

There are two obvious differences when looking at the diesel piston when compared to the petrol one. Firstly, there is no spark plug; the fuel–air mix is ignited in a different way. Secondly, the pistons are usually longer in a diesel engine for reasons that will become apparent.

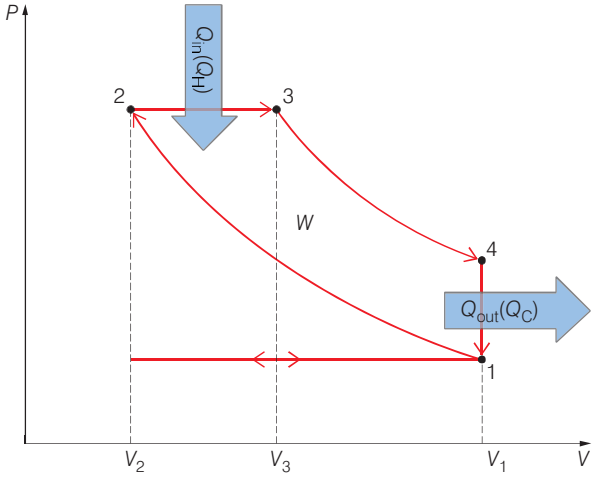


Figure 1.40 *An idealised diesel cycle*

The diagram in Figure 1.40 shows an idealised p–V diagram for a diesel cycle.

Again starting on the bottom left of the diagram the volume of the gas inside the piston is V_2. The piston is pulled down (due to the rotation of the shaft), air is drawn in through an open inlet valve but no fuel is present at this stage. The pressure remains constant and the volume of gas inside the cylinder increases to V_1.

The valve is closed and the piston moves back up, compressing the gas back to V_2. This process happens quickly and so the compression is adiabatic, no heat flows out of the gas. This can be seen in the line 1–2. As the piston is longer the air inside the piston is compressed much more than inside a petrol piston. This results in more work done on the gas and so a much larger increase in temperature.

At 2, diesel is squirted into the piston from the injector. The air is so hot the diesel ignites and begins to push the piston back down. The volume increases to V_3 but the pressure remains constant (isobaric). There is a significant amount of heat realised in this process (Q_{in} or Q_H).

At 3, all the diesel has combusted and work is done by the gas as it continues to force this piston down. This again happens very quickly without any heat flowing out of the system and so it is an adiabatic change. The volume of the gas increases back to V_1 and the pressure drops.

The rest of the process is the same as the petrol engine. Heat then flows out of the system (Q_{out} or Q_c) and the temperature of the gas drops causing a drop in pressure back to the original pressure. This can be seen in the line 4–1. The piston then moves back up with the exhaust valve is open and so the pressure remains constant as the waste gases are expelled. We are now back at the bottom left and the process starts again.

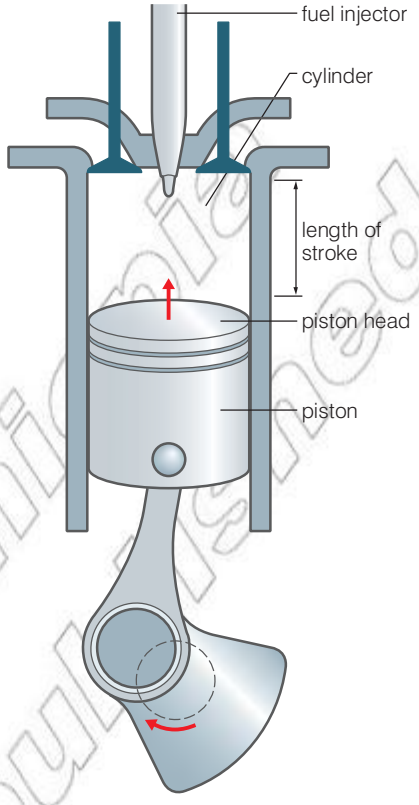


Figure 1.39 *A simple piston in a diesel engine*

© MDE EDR Ethiopia Not to be reproduced

The key difference between a diesel engine and a petrol one is the process for adding heat to the system. In a diesel engine it is isobaric, whereas in a petrol engine it is isochoric.

Other examples

There are many other examples of cycles employed by heat engines; these are summarised in Table 1.5. It is important to note there are two main groups.

The first group involves the burning of the fuel inside the working piston – internal combustion.

The second group involves burning a fuel in a chamber. The heat from this process is then transferred to a working fluid (such as steam) via a heat exchanger. This fluid is then transferred to the piston. As the heating does not take place inside the piston this is referred to as external combustion.

***Table 1.5** The various engine cycles*

Cycle	Compression of gas	Heat input	Expansion of gas	Heat output	Combustion
Otto (petrol engines)	adiabatic	isochoric	adiabatic	isochoric	internal
Diesel	adiabatic	isobaric	adiabatic	isochoric	internal
Brayton (jet engines)	adiabatic	isobaric	adiabatic	isobaric	internal
Carnot	adiabatic	isothermal	adiabatic	isothermal	external
Stirling	isothermal	isochoric	isothermal	isochoric	external
Ericsson	isothermal	isochoric	isothermal	isobaric	external
Rankine	adiabatic	isochoric	adiabatic	isobaric	external

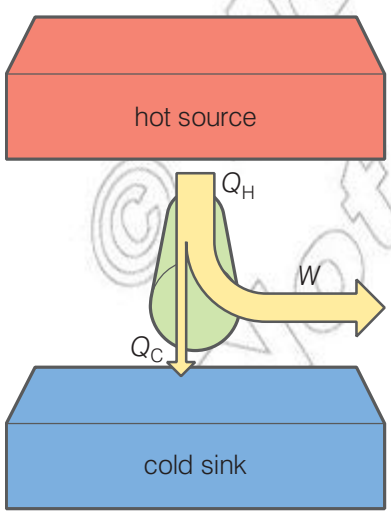


***Figure 1.41** The energies involved in a heat engine*

Efficiency of a heat engine

Just like any machine the efficiency of a heat engine may be determined by considering the energy input and the useful energy output.

- efficiency = useful work out / total work in

The diagram in Figure 1.41 shows the energies involved in a heat engine. The energy into the engine is Q_H and the useful work is W. Therefore the efficiency may be calculated using:

- efficiency = W / Q_H

By considering the law of conservation of energy we can express the efficiency in terms of Q_c and Q_H.

- energy from hot source = work done + energy to cold sink
- $Q_H = W + Q_c$

Therefore the useful work out of the engine may be found by:

- $W = Q_H - Q_c$

Applying the equation for efficiency we get:

- efficiency = useful work out / total work in
- $\eta = W / Q_H$
- $\eta = (Q_H - Q_c) / Q_H$

This may be expressed as:

- $\eta = 1 - (Q_c / Q_H)$

What is the maximum theoretical efficiency of a heat engine?

In order for heat energy to be drawn into the heat engine there must be temperature difference between the hot source and the cold sink. Some energy must always flow into the cold sink. This means Q_c can never be zero.

As a consequence the efficiency can never be 1 (i.e. 100% efficient).

As already discussed, this is confirmed by the second law of thermodynamics, which may be expressed as:

- **The complete conversion of energy from a hot source into work is not possible.**

However, the greater the absolute temperature difference between the hot source and cold sink, the greater the maximum theoretical efficiency.

- $\eta_{max} = 1 - (T_c / T_H)$

For example, the temperature inside a diesel engine may be 500 °C or 773 K. In this case the cold sink is the coolant in the engine. Assuming this is at 100 °C (373 K) the maximum theoretical efficiency will be:

- $\eta_{max} = 1 - (T_c / T_H)$
- $\eta_{max} = 1 - (373 / 773)$
- $\eta_{max} = 0.52$

However, no diesel engine is this efficient. A more common value is around 30% efficient (petrol engines have an even lower efficiency). The design of the engine results in several energy losses, mainly through friction. This results in additional heat being transferred to the surroundings.

However, the general principle is true that if the temperature difference between the hot source and cold sink is larger then the maximum theoretical efficiency increases.

In practice, increasing the temperature of the hot source causes significant problems in terms the material expansion of the components inside the engine. Engine designers are continuously researching new materials that can withstand higher temperatures in order to increase the efficiency of future engine designs.

Worked example 1.6

The heat input to a heat engine is 100 kJ. 25 kJ enters the cold sink. Determine the efficiency of the heat engine.

- $\eta = 1 - (Q_c / Q_H)$
- $\eta = 1 - (25\,000/100\,000)$
- $\eta = 0.75$ or 75%

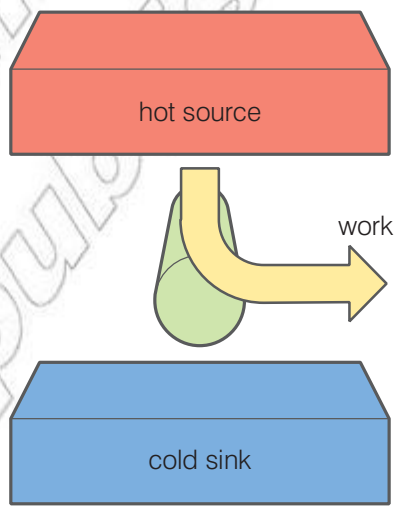


Figure 1.42 *Converting all the heat to work is forbidden by the second law.*

© MoE, FDR Ethiopia
Not to be republished

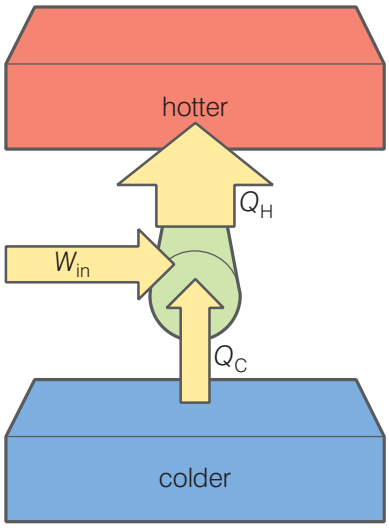


Figure 1.43 *A schematic of a heat pump*



Figure 1.44 *A refrigerator is a kind of heat pump.*

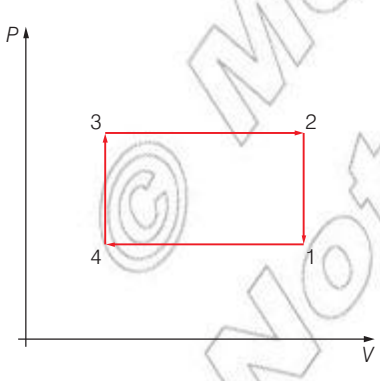


Figure 1.45 *A simple p–V diagram for a heat pump*

Refrigerators and heat pumps

Another way of stating the second law of thermodynamics is:

- **The spontaneous transfer of energy from a cooler body to a hotter body is not possible.**

This statement is most definitely true, in which case how it is possible for refrigerators to function? A bottle of water at room temperature goes into a cooler fridge and thermal energy is extracted from the water and transferred back to the room – which is at a higher temperature than the inside of the fridge.

It appears that the energy is flowing from a colder region to a hotter one, and it is! The key phrase here to consider is 'spontaneous transfer'. This means it does not happen without an additional input of energy.

A refrigerator is a kind of heat pump. A heat pump can be thought of as a heat engine running in reverse.

Work goes into the heat pump and this allows energy to flow from a colder region to a hotter one.

Imagine running the simple four-step cycle we looked at in heat engines running in reverse.

In this case there is more work done on the system than work done by the system. This net input of work (W_{in}) makes it possible for energy to be transferred from a colder region to a hotter one.

The energy transferred to the hotter region is given by:

- energy into hotter region = work input + energy from colder region
- $Q_H = W_{in} + Q_c$

How do refrigerators work?

Figure 1.46 shows a simple refrigerator. Inside every fridge there is a network of pipes containing a special fluid. The movement of this fluid around the fridge transfers thermal energy from inside the fridge (the colder region) to the surroundings (the hotter region).

Starting at the top left the liquid passes through an expansion valve. This is specially designed to force the liquid to change state into a gas and expand. This results in the gas doing work and so its temperature falls. The gas is pumped around the system, through the inside of the fridge.

Thermal energy flows from the inside of the fridge into the gas. This results in the temperature inside the fridge falling. This does not break the second law as the gas is much colder than the inside of the fridge.

Work is then done on the gas in order to compress it and turn it back into a liquid. This is the work put into the system (W_{in}). This causes an increase in temperature inside the fluid.

© MDR Education Not to be republished

The fluid then passes through a series of pipes on the back of the refrigerator. As it is now hotter than the surroundings and so thermal energy is transferred to the room (Q_H).

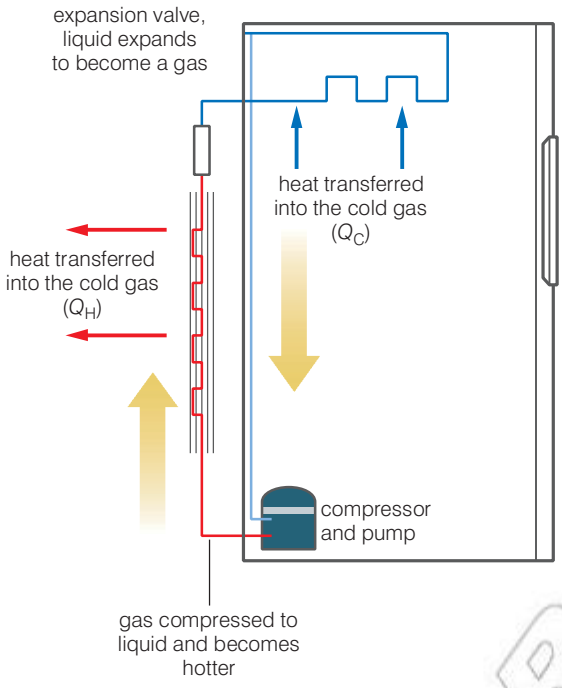


***Figure 1.46** A simple refrigerator*

Just like a heat engine, this process then cycles. The liquid passes through the expansion valve, turns into a gas and cools.

Think about this...

What would happen to the temperature of a room if a refrigerator was left running with the door open?

Activity 1.9: Refrigerators

Sketch a simple p–V diagram for a refrigerator, labelling each section.

Every cycle results in energy being transferred from the colder region to the hotter one. The food stays cold, despite heat leaking in from the surroundings, most often when you open the door!

DID YOU KNOW?

Other examples of heat pumps include air conditioners and freezers.

Summary

In this section you have learnt that:

- A heat engine converts some of the energy flowing from a hot source to a cold sink into mechanical work.
- Heat engines employ various thermodynamic processes to do mechanical work.
- The efficiency of a heat engine is given by $\eta = 1 - (Q_c / Q_H)$.
- All heat engines 'waste' energy in the form of heat transferred to their surroundings.
- The maximum theoretical efficiency of a heat engine is given by $\eta_{max} = 1 - (T_c / T_H)$.
- A heat pump allows thermal energy to be transferred from a colder region to a hotter one. This requires a net input of work.

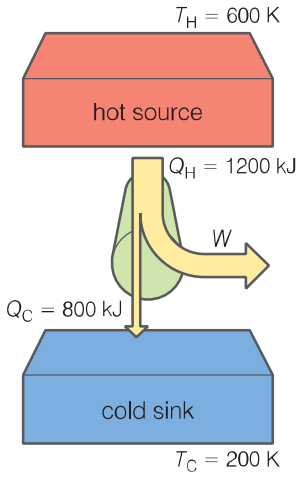


***Figure 1.47** For question 2*

Review questions

1. Describe what is meant by a heat engine.
2. Calculate the efficiency and the maximum theoretical efficiency of the heat engine in Figure 1.47.
3. Define the terms isochoric, adiabatic, isothermal and isobaric. Include a p–V diagram for each.
4. With the aid of a p–V diagram show how a net output of work may be produced from a cycle of thermodynamic processes
5. With reference to the second law of thermodynamics explain how refrigerators allow heat to be transferred from a colder region to a hotter one. Include a simple energy flow diagram with all key features labelled.

End of unit questions

1. The escape velocity from the Earth is approximately 11 km/s. This means that a gas molecule at the top of the atmosphere travelling outwards at 11 km/s will escape.
   a) The thermosphere, the layer of the upper atmosphere in which the Space Shuttle orbits, is at a temperature of about 1000 K. Calculate the mean kinetic energy of a molecule at this temperature.
   b) Calculate the r.m.s. speeds of (i) hydrogen, with molar mass 0.0020 kg/mol, and (ii) helium, with molar mass 0.040 kg/mol, at this temperature.
   c) If the r.m.s. speed of the molecules of a gas is greater than 0.2 of the escape velocity, then over the period of the Earth's existence all of the gas will have escaped from the atmosphere. Use this fact to explain whether we expect to find any hydrogen of helium in the atmosphere.
2. A meteorological balloon rises through the atmosphere until it expands to a volume of 1.0×10^8 m^3, where the pressure is 1.0×10^3 Pa. The temperature also falls from 17 °C to –43 °C. The pressure of the atmosphere at the Earth's surface is 1.0×10^5 Pa.
   a) What was the volume of the balloon at take-off?
   b) Before being released, the balloon was filled with helium gas of molar mass 4.0×10^{-3} kg/mol. Calculate (i) the number of moles of gas in the balloon, and (ii) the mass of gas in the balloon.
   c) When the balloon is filled at ground level, the internal energy of the gas is 1900 MJ. If the internal energy of the helium gas is equal to the random kinetic energy of all its molecules, estimate the internal energy of the helium when the balloon has risen to a height where the temperature is –43 °C.

© MoE, FDRE Ethiopia
Not to be republished

3. a) Explain what is meant by *internal energy*. Hence suggest how the internal energy of a *real* gas differs from that of an *ideal* gas.

   b) Figure 1.48 shows the graph of the cooling curve of a substance between 250 °C and room temperature. Over which sections of the curve is:

      i) the internal energy of the substance decreasing?

      ii) the average random kinetic energy of the molecules decreasing?

      iii) the average random potential energy of the molecules almost constant?

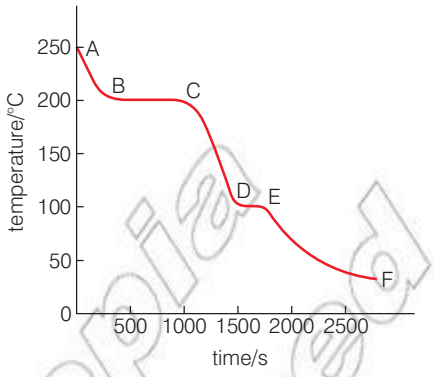


Figure 1.48

4. For every 10 m you descend below the surface of water, the pressure on you will increase by an amount equal to atmospheric pressure, 101 kPa. An air bubble in a lake has a volume of 20 mm^3 at a depth of 40 m. Predict what volume it will have just before it reaches the surface and write down the assumption that you have made in order to arrive at an answer.

5. Figure 1.49 shows curves (not to scale) relating pressure, p, and volume, V, for a fixed mass of an ideal monatomic gas at 300 K and 500 K. The gas is in a container fitted with a piston which can move with negligible friction.

   a) Show that the number of moles of gas in the container is 0.0201.

   b) Show that the volume of the gas at B on the graph is 1.67×10^{-3} m^3.

   c) Calculate the total internal energy of the gas in the container at point A on the graph.

   d) Explain how the first law of thermodynamics applies to the changes represented on the graph by (i) A to C, and (ii) A to B. Calculate the energy absorbed in each case by heating.

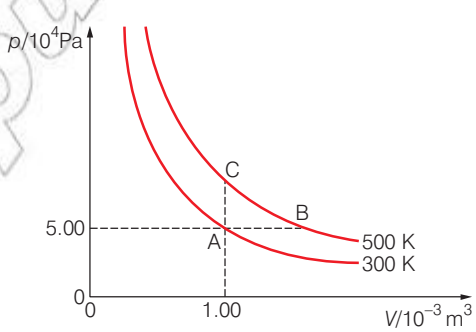


Figure 1.49

6. a) A perpetual motion machine would be able to produce a continuous output of work with no energy input. State the physical principle that makes this impossible.

   b) Figure 1.50 shows one suggestion put forward as a perpetual motion machine. The ball in position **A** would fall off the top of the water doing work on the pulley belt. At **B** it would move sideways doing no work and enter the bottom of the tank by a valve system which would prevent water from escaping. It would then float to the top ready to start again. Explain why this system will not behave as a perpetual motion machine.

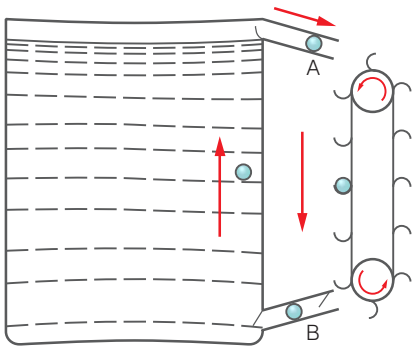


Figure 1.50

7. By tidying up your house last night, you will have decreased the amount of disorder present. In order for the second law of thermodynamics to still apply what did you do to the air in your house whilst you were tidying?

© Mofed FDR Ethiopia Not to be republished

8. Sketch the p–V diagram for a petrol engine and describe each of the thermodynamic processes involved.
9. A smoke jack is another type of heat engine used to cook meat. Heat from a fire causes air to rise above the fire. This rising air causes a paddle wheel in the chimney to turn. This in turn rotates an animal being roasted above the fire.
   a) The temperature of the surrounding air is 15 °C and the fire is at a temperature of 350 °C. Calculate the maximum theoretical efficiency of the smoke jack.
   b) Sketch a simple diagram of a smoke jack and use this to indentify two places where the energy would be lost.
10. A heat engine operating between 100 °C and 700 °C has an efficiency equal to 40% of the maximum theoretical efficiency. How much energy does this engine extract from the hot reservoir in order to do 5000 J of mechanical work?

© MoE, FDR Ethiopia
Not to be republished

Oscillations and waves

Unit 2

Contents

Section	Learning competencies
2.1 Periodic motion (basic concepts) (page 53)	• Describe the periodic motion of a vibrating object in qualitative terms, and analyse it in quantitative terms (e.g. the motion of a pendulum, a vibrating spring, a tuning fork). • Define simple harmonic motion (SHM) and describe the relationship between SHM and circular motion. • Derive and use expressions for the frequency, periodic time, displacement, velocity and acceleration of objects performing SHM. • Draw and analyse x–t, v–t and a–t graphs for SHM. • Use Newton's second law and Hooke's law to derive $\omega = \sqrt{k/m}$. • Describe the effects: free oscillations, damping, forced oscillations and resonance. • Analyse the components of resonance and identify the conditions required for resonance to occur in vibrating objects and in various media, including the effects of damping on resonance. • Explain the energy changes that occur when a body performs SHM. • Draw and interpret graphs showing the variation of kinetic energy and potential energy of an object performing SHM. • Relate the energy of an oscillator to its amplitude. • Solve problems on SHM involving period of vibration and energy transfer.
2.2 Wave motion (page 80)	• Describe the characteristics of a mechanical wave and identify that the speed of the wave depends on the nature of medium. • Use the equation $v = \sqrt{T/\mu}$ to solve related problems • Describe the characteristics of a travelling wave and derive the standard equation $y = A\cos(\omega t + \phi)$ • Define the terms phase, phase speed and phase constant for a travelling wave. • Explain and graphically illustrate the principle of superposition, and identify examples of constructive and destructive interference. • Identify the properties of standing waves and for both mechanical and sound waves, explain the conditions for standing waves to occur, including definitions of the terms node and antinode. • Derive the standing wave equations. • Calculate the frequency of the harmonics along a string, an open pipe and a pipe closed at one end. • Explain the modes of vibration of strings and solve problems involving vibrating strings. • Explain the way air columns vibrate and solve problems involving vibrating air columns.

Contents

Section	Learning competencies
	• Analyse, in quantitative terms, the conditions needed for resonance in air columns, and explain how resonance is used in a variety of situations. • Identify musical instruments using air columns, and explain how different notes are produced.
2.3 Sound, loudness and the human ear (page 97)	• Define the intensity of sound and state the relationship between intensity and distance from the source. • Describe the dependence of the speed of sound on the bulk modulus and density of the medium. Use $v = \sqrt{B/\rho}$ • Give intensity of sound in decibels, and define the terms threshold of pain and threshold of hearing. • Describe the intensity level versus frequency graph to know which the human ear is most sensitive to. • Explain the Doppler effect, and predict in qualitative terms the frequency change that will occur in a variety of conditions. • Explain some practical applications of the Doppler effect.

KEY WORDS

simple harmonic motion *the periodic oscillation of an object about an equilibrium position, such that its acceleration is always directly proportional in size but opposite in direction to its displacement*

oscillating *vibrating about a central position*

equilibrium position *the position of an oscillating object when at rest*

restoring force *the force on a displaced object that acts towards its original position*

A great many things in the world around us oscillate (vibrate) backwards and forwards, up and down, side to side, in and out, etc. Atoms within molecules vibrate and the size of these vibrations is proportional to temperature. Oscillations of charges produce electromagnetic waves: e.g. a current oscillating up and down an aerial produces radio waves. Vibrations of our vocal chords produce sound waves, as do vibrations of strings and of air inside tubes in musical instruments. Parts of machinery, e.g. in washing machines and in cars, vibrate, sometimes when we don't want them to!

When engineers build large structures like skyscrapers and bridges, they have to understand how the wind or the ways people walk across them will make them oscillate. It is impossible to stop such structures oscillating altogether, but if engineers don't design their structures to control these vibrations, they might end up shaking themselves to pieces.

Most of these oscillations are periodic. This means that they keep doing exactly the same thing in the same amount of time again and again. In some cases, usually for large objects or structures, each cycle (backwards and forwards, up and down, side to side, in and out, etc.) of the oscillation could take many seconds or even much longer. These are low-frequency oscillations. In other cases there can be hundreds, thousands or even thousands of billions of complete vibrations every second. These are high-frequency oscillations. This predictable time period can be very useful. For example, the predictable time period of pendulums, of masses on springs or of quartz crystals is used to count the passing of time in clocks and watches.

The way things oscillate can be quite complex, but many oscillations are very close to a special form of periodic motion

© NOT FOR Not to be copied

called **simple harmonic motion**, and more complicated motion can be shown to be simply a sum of simple harmonic motions at different frequencies. This unit will analyse a few examples of oscillating objects performing simple harmonic motion in some mathematical detail.

2.1 Periodic motion (basic concepts)

By the end of this section you should be able to:

- Describe the periodic motion of a vibrating object in qualitative terms, and analyse it in quantitative terms (e.g. the motion of a pendulum, a vibrating spring, a tuning fork).
- Define simple harmonic motion (SHM) and describe the relationship between SHM and circular motion.
- Derive and use expressions for the frequency, periodic time, displacement, velocity and acceleration of objects performing SHM.
- Draw and analyse x–t, v–t and a–t graphs for SHM.
- Use Newton's second law and Hooke's law to derive $\omega = \sqrt{k/m}$.
- Describe the effects: free oscillations, damping, forced oscillations and resonance.
- Analyse the components of resonance and identify the conditions required for resonance to occur in vibrating objects and in various media, including the effects of damping on resonance.
- Explain the energy changes that occur when a body performs SHM.
- Draw and interpret graphs showing the variation of kinetic energy and potential energy of an object performing SHM.
- Relate the energy of an oscillator to its amplitude.
- Solve problems on SHM involving period of vibration and energy transfer.

Periodic oscillations

If something is **oscillating** (vibrating) this means that it is moving backwards and forwards, up and down, side to side, in and out, etc, around some central position. This central position is called the equilibrium position and it is the position of the object when it is at rest.

Whenever an object is displaced from its **equilibrium position** there is a force that acts towards its original position. This force is often referred to as a **restoring force**, as it tries to restore the system to its equilibrium position. This is much easier to understand if we look at some simple examples.

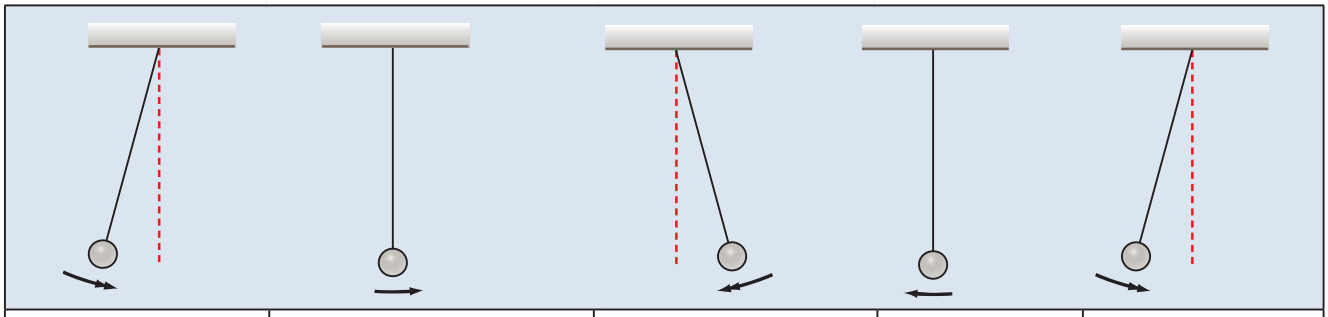
DID YOU KNOW?

The pendulum clock was invented in 1656 by Dutch scientist Christiaan Huygens. Huygens was inspired by investigations of pendulums by Galilei Galileo, beginning around 1602. Galileo discovered the key property that makes pendulums useful timekeepers: isochronism, which means that the period of swing of a pendulum is approximately the same for different sized swings. Up until the 1930s, the pendulum clock was the world's most accurate timekeeper, but they must be stationary to operate as any motion or accelerations will affect the motion of the pendulum, causing inaccuracies, and so they could never be used for portable devices. They are now out of date of course; we now have more accurate devices, though still using simple harmonic motion.



Figure 2.1 *A simple pendulum clock*

How does a pendulum work?



(a)	(b)	(c)	(d)	(e)
If the pendulum bob is pulled to one side and released, it accelerates back towards its equilibrium position.	When the pendulum bob gets back to the equilibrium position, it is moving relatively fast and, although there is no resultant force now, its inertia keeps it moving.	The pendulum bob keeps moving, slowing down all the time, until it is at the same height as it started and accelerates back towards the equilibrium position.	The pendulum bob passes through the equilibrium position again, going back the other way.	The pendulum bob arrives back at where it started. It has completed one cycle, and will now do the same again, and again . . .

KEY WORDS

resultant force *the overall force acting on an object*

acceleration *rate of change of velocity*

Figure 2.2 *Oscillation of a pendulum when the bob is pulled to one side and released*

A simple pendulum is made by hanging a mass, known as the bob, on a string from a fixed support, as shown in Figure 2.2.

If we let the mass hang without swinging, it will hang directly below the support with all forces on it balanced. This position, where the **resultant force** acting on the bob is zero, is known as the equilibrium position.

If we give the bob a small initial displacement by pulling it to one side and then release it, there will be a resultant force, due to the weight of the bob and the tension acting in the string. This force pulls it back towards the equilibrium position. This causes **acceleration** towards the equilibrium position (opposite to the direction of displacement).

When the bob reaches the equilibrium position, the resultant force is now zero, but the bob is moving and can't stop instantly. Its inertia keeps it moving through the equilibrium position, and if there is no significant friction of air resistance, it will keep moving, slowing down all the time until it is as high as it was when it started.

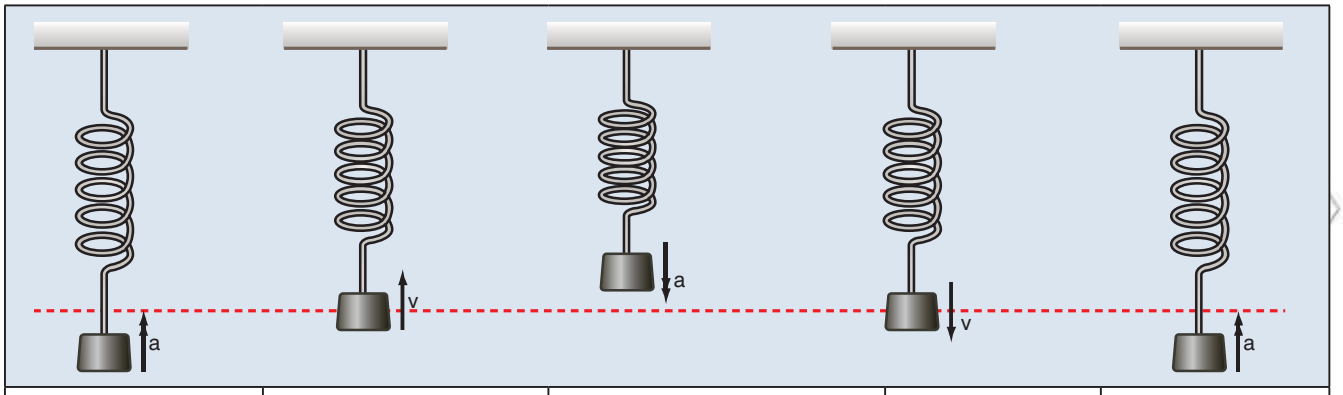
It now has a displacement equal and opposite to its starting displacement. However, as displacement is a vector quantity it is now a negative value. If the initial displacement was 3 cm, the displacement after one swing (half an oscillation) will be –3 cm.

In exactly the same way, it will swing back to where it started to complete one complete cycle of the oscillation. It will now repeat this process again and again.

It is important to notice the force causing the oscillation always acts towards the equilibrium position.

© MoE TBK Not to be resold

How does a mass on a spring oscillate?



(a)	(b)	(c)	(d)	(e)
If the mass is pulled down, the tension in the spring is greater the weight of the mass. The resultant force is upwards and the mass accelerates upwards when released.	When the mass gets back to the equilibrium position, it is moving relatively fast and although the weight of the mass balances the tension in the spring and there is no resultant force now, its inertia keeps it moving.	The mass keeps moving, slowing down all the time, until it is as far above the equilibrium point as it started below. The weight of the mass is greater than the tension in the spring, the resultant force is now downwards and the mass accelerates downwards.	The mass passes through the equilibrium position again, going back the other way.	The mass arrives back at where it started. It has completed one cycle, and will now do the same again, and again . . .

Figure 2.3 *Oscillation of a mass–spring system when the mass is displaced downwards and released*

If a mass is hung from a support by a spring and allowed to settle until it is stationary, it will hang with the spring stretched so that the restoring force (in this case the tension in the spring) is equal and opposite to the weight of the mass. This is the equilibrium position.

If we now pull the mass down, the tension in the spring will be greater than the weight of the mass. The resultant force on the mass is upwards and so, if we let go, it accelerates upwards. When the mass gets back to the equilibrium position it is moving and, although there is no resultant force here, its inertia keeps it moving.

The mass keeps moving, slowing down all the time, until it is as far above the equilibrium point as it started below. The tension in the spring is now less than the weight of the mass, the resultant force is now downwards and the mass accelerates downwards. The mass passes through the equilibrium position again, and carries on until it arrives back at where it started. It has completed one cycle, and will now do the same again, and again …

© MOED Not to be re

Activity 2.1: Mass–spring forces

Sketch a diagram of the forces acting on a mass–spring system:

a) when in equilibrium position

b) at the bottom of its oscillation

c) at the top of its oscillation.

This process would also happen if the spring was horizontal on a low friction surface.

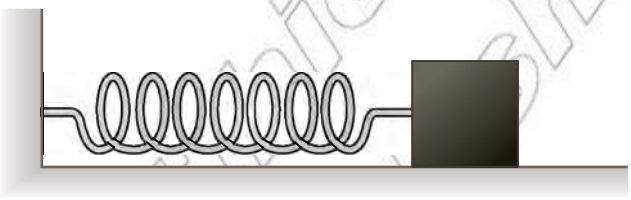


Figure 2.4 *A horizontal mass–spring system*



Figure 2.5 *Vehicle suspensions can act like mass–spring systems.*

Discussion activity

We have used the words displacement and acceleration in describing the motions of the pendulum and the mass–spring system. These are vector quantities: their directions are very important. When the bob or mass is moving away from the equilibrium position and slowing down, which direction is the acceleration in? What can you say about the direction of the acceleration (i) relative to the equilibrium position, and (ii) relative to the direction of displacement? What do you know in general about the acceleration of an object and the resultant force acting on it?

How do we define SHM?

SHM

- **The acceleration is proportional to the displacement.**
- **The acceleration is in the opposite direction to the displacement.**

Simple harmonic motion (SHM) is a periodic oscillation of an object about an equilibrium position such that its acceleration is always directly proportional in size but opposite in direction to its displacement. (The acceleration is always towards the equilibrium position.)

This defining relationship is shown in Figure 2.6. This graph is much simpler than many graphs that will follow later in this unit, but **it is the most important**.

It follows from Hooke's law that the restoring force has the same relationship to the displacement (as forces and acceleration are directly proportional). The greater the displacement from the equilibrium position the greater the restoring force, and this force acts in the opposite direction to the displacement.

Consider the mass–spring system. When the spring is most extended, it is furthest from its equilibrium position. At this point the restoring force is also at its greatest, but it acts in the opposite direction.

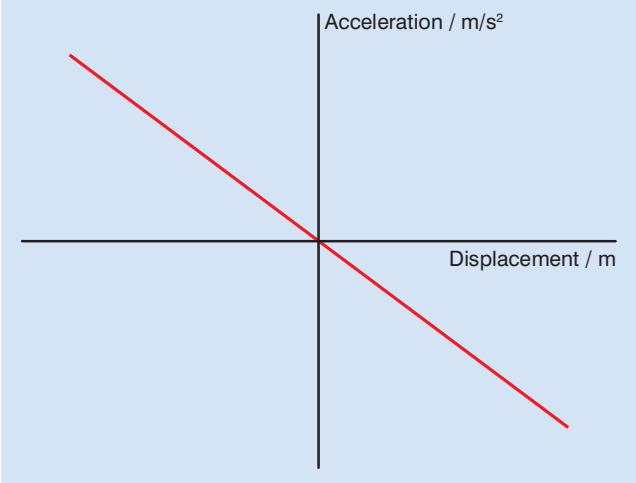


Figure 2.6 *Defining relationship for SHM: acceleration is directly proportional but opposite in sign to displacement.*

What does SHM look like?

If we plot how the **displacement** of an object performing simple harmonic motion varies with time, we find that the variation is **sinusoidal**, as shown in Figure 2.7. Note that the displacement goes positive and negative as the mass oscillates either side of the equilibrium position.

The size of the maximum displacement in either direction is called the **amplitude** *A*. The time to perform one complete cycle of the oscillation is called the **time period** *T*.

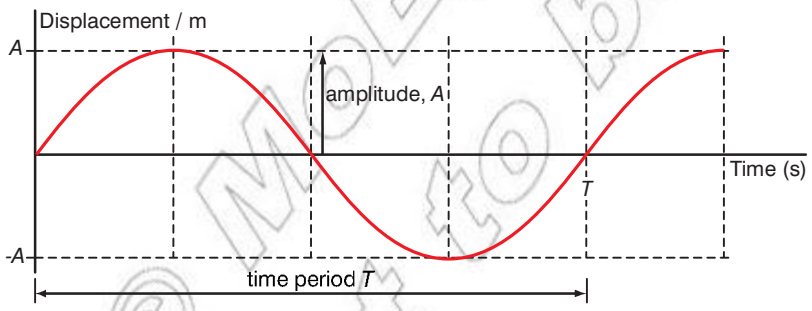


Figure 2.7 *Variation of displacement with time for simple harmonic motion*

When we say the oscillation is sinusoidal, we mean that the displacement is described mathematically using sine or cosine functions:

$$x = A \sin\left(2\pi \frac{t}{T}\right) \text{ or } x = A \cos\left(2\pi \frac{t}{T}\right),$$

KEY WORDS

displacement *the distance moved in a specific direction*

sinusoidal *an oscillation that can be described mathematically using sine or cosine functions*

amplitude *the maximum displacement of the wave from the equilibrium position*

time period *the time taken for one complete cycle of an oscillation*

© MoE FDRE Ethiopia Not to be republished

where A is the amplitude of the oscillation and T the time period. Either could be used, but throughout the rest of this chapter we will use,

$$x = A \sin\left(2\pi \frac{t}{T}\right),$$

although the cosine function gives a better description if the SHM is started by displacing the oscillator and then releasing it.

If $x = A \sin\left(2\pi \frac{t}{T}\right)$, with $\left(\frac{2\pi}{T}t\right)$ expressed in radians:

when	$t = 0$	$x = A \sin\left(\frac{2\pi}{T}0\right)$	$= A \sin(0)$	$= 0$
	$t = \frac{T}{4}$	$x = A \sin\left(\frac{2\pi}{T}\frac{T}{4}\right)$	$= A \sin\left(\frac{\pi}{2}\right)$	$= A$
	$t = \frac{T}{2}$	$x = A \sin\left(\frac{2\pi}{T}\frac{T}{2}\right)$	$= A \sin(\pi)$	$= 0$
	$t = \frac{3T}{4}$	$x = A \sin\left(\frac{2\pi}{T}\frac{3T}{4}\right)$	$= A \sin\left(\frac{3\pi}{2}\right)$	$= -A$
	$t = T$	$x = A \sin\left(\frac{2\pi}{T}T\right)$	$= A \sin(2\pi)$	$= 0$

Looking carefully at the information above you can see how in one oscillation the displacement starts at 0 rises to a positive amplitude, falls back to zero, falls to a negative amplitude and then rises back to zero.

Activity 2.2: Displacement using cosine

Use the same method above to show how the displacement varies if cosine were to be used instead of sine. Sketch the corresponding displacement–time graph.

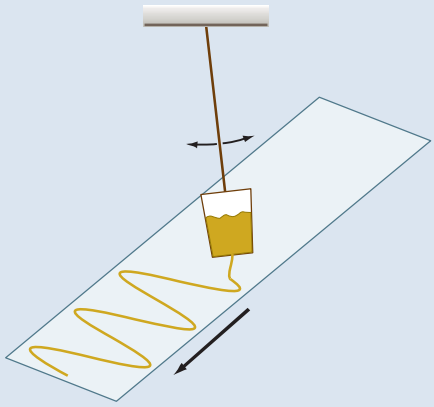
Discussion activity

A sinusoidal motion looks fairly complicated, so why is simple harmonic motion called simple? Looking at Figure 2.7 should give you a clue.

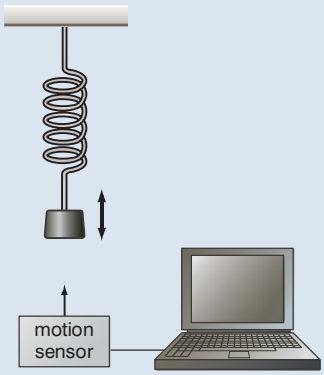
© MoE-FDRE Ethiopia Not to be republished

How can we observe SHM?

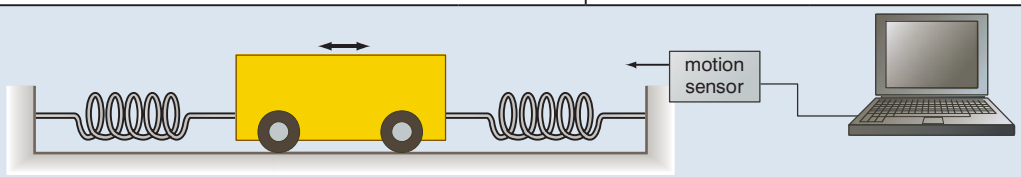
Figure 2.8 shows a number of ways of obtaining a graph of displacement against time for oscillators performing SHM.



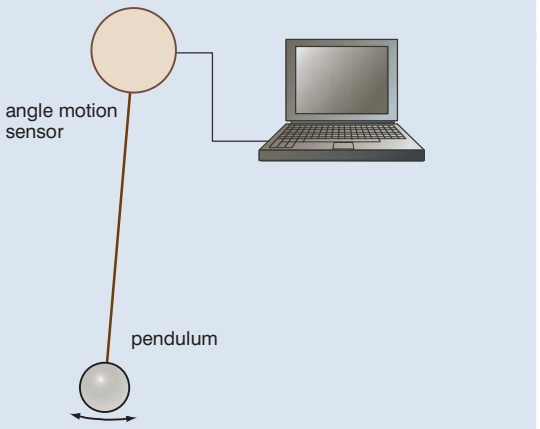
Hang a small bucket with a hole in the bottom on a rope. Fill the bucket with sand so that a stream of sand runs out onto a long sheet of paper underneath. Start the bucket swinging and pull the paper at a constant speed and the sand will draw a sinusoidal wave.



If you have a motion sensor connected to a PC, there are several ways to record variation of displacement with time. One way is to place the sensor facing upwards underneath a mass hanging on a spring.



A dynamics trolley moving backwards and forwards with a anchored spring attached to both ends can make a more consistent target for a motion sensor.



An angular motion sensor is an easy way of observing how a pendulum swings.

Figure 2.8 *Experiments to observe SHM*

Think about this...

If the frequency of a mass–spring system is 50 Hz how many times in 1 second will the mass pass through its equilibrium position?

Activity 2.3: Displacements

For the same pendulum calculate the displacement after:

a) 1.2 s

b) 3.4 s

Explain the significance of the negative value.

Think about this...

It is a good idea to avoid using the word 'fast' when describing oscillations. Even if the frequency is low, if the amplitude of the oscillation is large, the oscillator will be moving quickly when it goes through the equilibrium position, and so the word fast could still apply. This can lead to confusion.

To give a good clear scientific description, simply talk about high or low frequencies and large or small amplitudes.

KEY WORDS

frequency *the number of cycles per second*

Frequency and time period

The **frequency**, f, of an oscillation is the number of cycles it completes per second. The unit is the hertz, symbol Hz. A frequency of 50 Hz would correspond to 50 complete oscillations per second.

Frequency is related to time period by:

- $f = \frac{1}{T}$

and so our mathematical expression for displacement can be written as

- $x = A \sin(2\pi ft)$.

Worked example 2.1

A pendulum has a frequency of 4.0 Hz and amplitude of 5.0 cm. Determine the displacement after 4.6 seconds.

- $x = A \sin(2\pi ft)$. *State the equation of SHM*
- $x = 0.050 \sin(2\pi \times 4.0 \times 4.6)$ *Substitute the known values*
- $x = 0.029$ m or 2.9 cm. *Solve the equation and give the units*

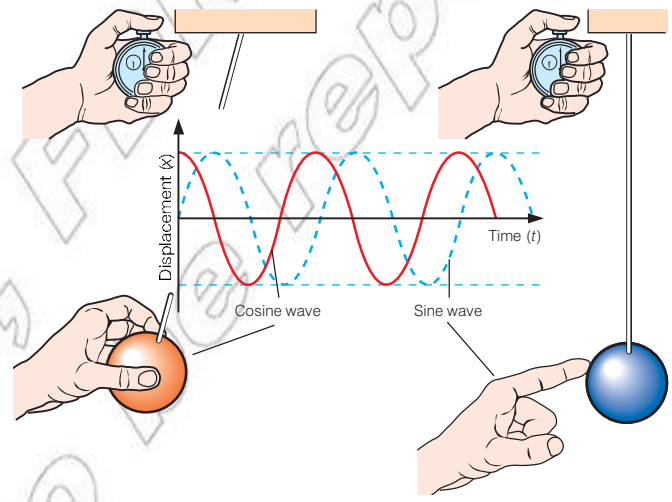


Figure 2.9 *Oscillations of an object*

Discussion activity

One feature of SHM, particularly useful in building clocks, is that, for perfect SHM, the frequency or time period does not vary with amplitude. So if an oscillator does lose energy and its amplitude fall over time, the time period will not change.

Circular motion, SHM and angular frequency

If a point P moves around in circle of radius A, as shown in Figure 2.10, starting from point C, then the height of point P, after it has turned through angle θ is given by

- $h = A\sin\theta$

We now need to introduce a new quantity: angular speed ω. Angular speed is the rate of change of angle turned θ with time, in exactly the same way that linear velocity v is the rate of change of linear displacement s with time. ω is measured in radians per second (rad/s).

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

If the point P is rotating at angular speed ω radians per seconds then, after time t seconds, the total angle turned, in radians, is

- $\theta = \omega t$

and so we can rewrite the equation for the height of point P as

- $h = A\sin(\omega t)$

If P goes round in one complete cycle, the angle turned is 2π. If P is rotating with a frequency of f cycles per second, the total angle turned per second is $f \times 2\pi$ radians. Hence

- $\omega = 2\pi f$

In the equation $x = A\sin(2\pi ft)$ we can replace $2\pi f$ by ω and write:

- $x = A\sin(\omega t)$

which is the same as our expression for the height of point P rotating at angular speed ω. The height h of point P going round in a circle and the displacement x of an object performing SHM are therefore the same, as shown in Figure 2.11.

Think about this...

We must use radians when considering all the equations of SHM. Remember 2π radians is equal to 360° or one complete oscillation.

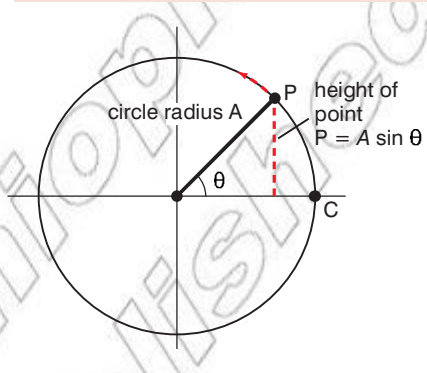


Figure 2.10 *The height h of point P when it has turned through angle θ from starting point C is given by* $h = A\sin(\omega t)$.

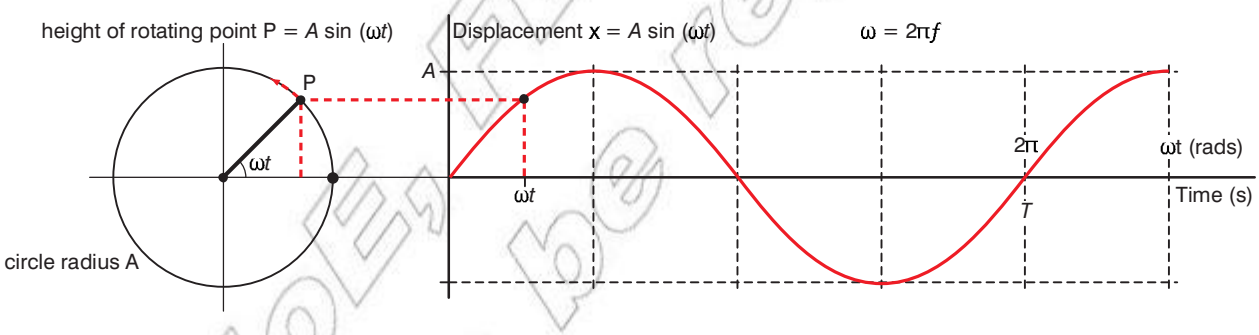


Figure 2.11 *After time t, the displacement x of an object performing SHM of amplitude A and frequency f is the same as the height h of a point P performing circular motion with radius A and angular speed* $\omega = 2\pi f$.

The relationship between **angular speed** and time period is

- $\omega = \frac{2\pi}{T}$

Because of its relationship to frequency, ω is sometimes called **angular frequency.**

KEY WORDS

angular speed *the rate of change of angle turned with time*

angular frequency *the rate of change of angular displacement*

© MoE-EDP Ethiopia Not to be reproduced

Discussion activity

Figure 2.12 shows a piston moving backwards and forwards inside a cylinder connected by a rod, hinged at both ends, to a rotating wheel.

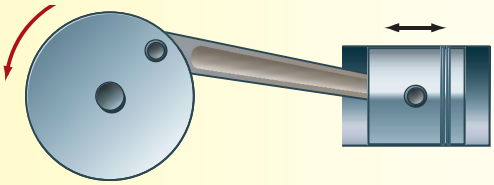


Figure 2.12 *Motion of a piston*

Under what conditions does the motion of the piston approximate to SHM?

KEY WORDS

rate of change *The speed at which a variable changes over a specific period of time*

Displacement, velocity and acceleration in SHM

The velocity of the oscillating mass is the **rate of change** of displacement. It becomes zero at the limits of the oscillation. For example, at the top of a pendulum's swing.

In general, velocity can be found as the gradient of the displacement–time graph. At the maximum displacement (the amplitude) the gradient is zero – consequently the velocity is zero. Therefore, when $x = \pm A$ the velocity is zero.

The maximum velocity occurs when the oscillating object passes through the equilibrium position. Again this may be seen on the displacement–time graph. The gradient of the line is at its greatest when it passes through the equilibrium position therefore the velocity is greatest at this point.

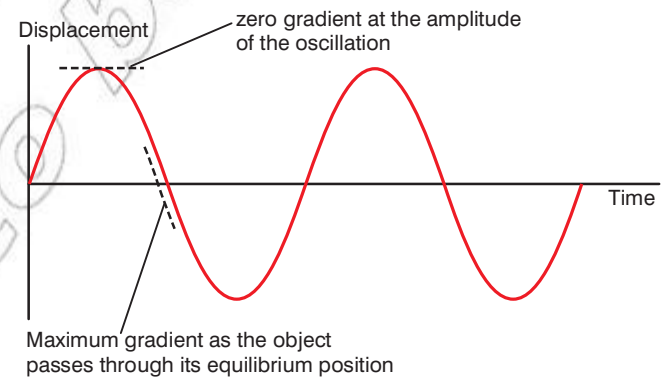


Figure 2.13 *The gradient of the displacement–time graph is equal to the velocity of the object.*

The equation of the oscillation shown in Figure 2.13 is $x = A\sin(\omega t)$. Therefore the velocity may be found by:

- $v = \omega A\cos(\omega t)$

© MoE; TRP; Not to be reproduced

and the maximum size of velocity (v_o)

- $v_o = \omega A$

This equation is obtained when $\cos(\omega t) = 1$. This happens whenever the mass passes through the equilibrium position.

Acceleration is the **rate of change** of velocity, or the gradient of the velocity time graph. This can be shown to give

- $\boldsymbol{a = -\omega^2 A \sin(\omega t)}$

and the maximum **size** of acceleration (a_o)

- $a_o = \omega^2 A$

This equation is obtained when $\sin(\omega t) = 1$. This happens whenever the mass reaches its maximum displacement.

Since $A\sin(\omega t) = x$, this is the same as

- $\boldsymbol{a = -\omega^2 A x}$

This is the defining equation for SHM.

We have already stated that acceleration is directly proportional and opposite in sign to displacement. We now see that the constant of proportionality is $-\omega^2$. Remember this is also equal to $(2\pi f)^2$ or $(2\pi/T)^2$.

A graph of acceleration plotted against time will look like an upside down version of the graph of displacement, emphasising the crucial point that acceleration is always in the opposite direction to displacement.

Graphs of displacement, velocity and acceleration against time t or angle ωt are shown overleaf in Figure 2.14.

Two key points to note and check for yourself looking at these is that:

- **the velocity at any time is the gradient of the displacement-time graph at that time and the acceleration at any time is the gradient of the velocity–time graph at that time, and**
- **the acceleration is directly proportional to and opposite in sign to the displacement.**

If the oscillator starts from the limit of oscillation at $x = A$, then displacement is better described using a cosine wave and the equations for displacement, velocity and acceleration become:

SHM equation summary

Remember you can use either the sine or cosine function to describe the displacement of a system oscillating with SHM. You need to consider when the timing of the oscillation begins.

- Timing starts with system in its equilibrium position → sine
- Timing starts with system at its maximum displacement → cosine

Think about this...

The equation for velocity can be found using differential calculus.

$$v = \frac{ds}{dt} = \frac{d}{dt}(A\sin(\omega t))$$
$$= \omega A\cos(\omega t)$$

Think about this...

The equation for acceleration can be found using differential calculus.

$$a = \frac{dv}{dt} = \frac{d}{dt}\omega A\cos(\omega t)$$
$$= -\omega^2 A\sin(\omega t)$$

The equation can be written as $a = -\omega^2 x$.

This is a differential equation and we have just shown that $x = A\sin(\omega t)$ is a solution of this differential equation. It can be shown that $x = A\cos(\omega t)$ is an equally valid solution, and that a more general solution is $x = A\cos(\omega t + \theta_o)$, where θ_o is an initial "phase" angle.

© MoE FDRE Ethiopia
Not to be republished

The equations are summarised below:

	Using sin to describe displacement	Using cos to describe displacement
Displacement	$x = A\sin(\omega t)$	$x = A\cos(\omega t)$
Velocity	$v = \omega A\cos(\omega t)$	$v = -\omega A\sin(\omega t)$
Acceleration	$a = -\omega^2 A\sin(\omega t)$	$a = -\omega^2 A\cos(\omega t)$

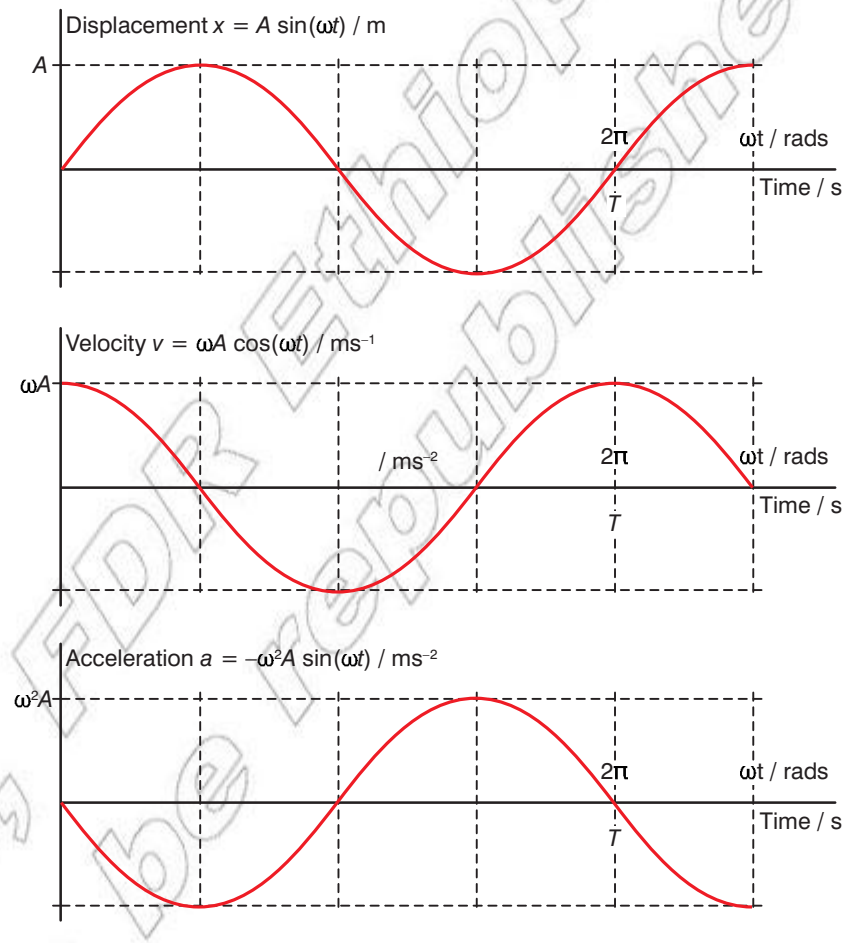


Figure 2.14 *Graphs of displacement, velocity and acceleration against time t or angle ωt for an object performing SHM*

Worked example 2.2

An object moves with simple harmonic motion of amplitude 11 cm and a time period of 2.4 s. Calculate:

a) the frequency

b) the angular frequency

c) the maximum velocity of the object

d) the maximum acceleration of the object

e) the displacement, velocity and acceleration, after 0.5 s, if the object starts from the limit of oscillation at $x = A$.

© MoE, FDR Ethiopia
Not to be republished

a) Frequency

- $f = \frac{1}{T}$ *State the relationship to be used*
- $f = \frac{1}{2.4} = 0.42$ Hz *Substitute in known values and solve, giving the units*

b) Angular frequency

- $\omega = 2\pi f = \frac{2\pi}{T}$ *State the relationship to be used*
- $\omega = \frac{2\pi}{2.4} = 2.6$ rad/s *Substitute in known values and solve, giving the units*

c) Maximum velocity

- $v_0 = \omega A$ *State the relationship to be used*
- $v_0 = \frac{2\pi \times 0.11}{2.4} = 0.29$ m/s *Substitute in known values and solve, giving the units*

d) Maximum acceleration

- $a = \omega^2 A$ *State the relationship to be used*
- $a = \left(\frac{2\pi}{2.4}\right)^2 \times 0.11 = 0.75$ m s^{-2} *Substitute in known values and solve, giving the units*

e) Using the cosine equation for displacement, after 0.5 s

- $x = A\cos(\omega t)$ *State the relationship to be used*
- $x = 0.11 \times \cos\left(\frac{2\pi}{2.4} \times 0.5\right)$
  $= 0.11 \times \cos(1.31)$
  $= 0.11 \times 0.259 = 0.028$ m

  Substitute in known values and solve, giving the units

The use the appropriate equation to find the velocity, in this case:

- $v = -\omega A\sin(\omega t)$
- $v = -\left(\frac{2\pi}{2.4}\right) \times 0.11 \times \sin(1.31)$
  $= -\left(\frac{2\pi}{2.4}\right) \times 0.11 \times 0.966$
  $= -0.28$ m s^{-1}

  Substitute in known values and solve, giving the units

Finally use the defining equation for SHM to find the acceleration.

- $a = -\omega^2 A\cos(\omega t)$ *State the relationship to be used*
- $a = -\left(\frac{2\pi}{2.4}\right)^2 \times 0.11 \times \cos\left(\frac{2\pi}{2.4} \times 0.5\right) = -0.20$ m s^{-2}

  Substitute in known values and solve, giving the units

How does velocity depend on displacement?

We already know that the velocity is zero when $x = \pm A$ at the limits of the oscillation, and that it has its maximum size of $v_o = \omega A$ when the mass passes through the equilibrium position.

To get a general expression we need to use a trigonometric identity:

- $\sin^2\theta + \cos^2\theta = 1$

  If $v = -\omega A\cos(\omega t)$ we have $v^2 = \omega^2 A^2\cos^2(\omega t)$

  If $x = A\sin(\omega t)$ we have $\omega^2 x^2 = \omega^2 A^2\sin^2(\omega t)$

and so

- $v^2 + \omega^2 x^2 = \omega^2 A^2\cos^2(\omega t) + \omega^2 A^2\sin^2(\omega t)$
- $v^2 + \omega^2 x^2 = \omega^2 A^2[\cos^2(\omega t) + \sin^2(\omega t)]$
- $v^2 + \omega^2 x^2 = \omega^2 A^2 \times 1$
- $v^2 + \omega^2 x^2 = \omega^2 A^2$
- $v^2 = \omega^2 A^2 - \omega^2 x^2$
- $v = \pm\omega\sqrt{A^2 - x^2}$

This equation shows us that at any given displacement (x) an oscillating object may have +/– a specific velocity. This is easy to explain.

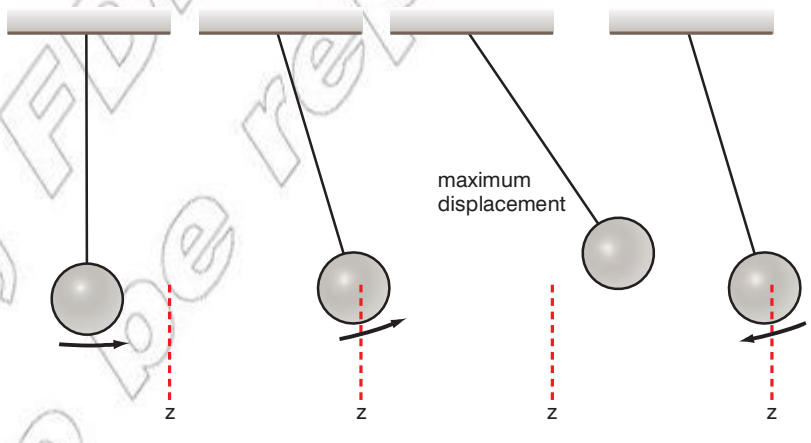


Figure 2.15 *As part of any oscillation the mass will pass through the same point twice.*

If you consider a simple pendulum swinging towards its maximum displacement, on its way up it passes through point Z. It then stops and swings back through Z in the opposite direction. Therefore at any given displacement a pendulum bob may have a velocity equal to $+v$ or $-v$.

If the displacement is 0 (i.e. as the mass passes through the equilibrium position) $x = 0$. Therefore $v = \pm\omega\sqrt{A^2 - x^2}$ becomes $v = \pm\omega\sqrt{A^2 - 0^2}$. This simplifies to $v_o = \pm\omega A$.

© MoE, FDRE Ethiopia
Not to be republished

How do we calculate the time periods of real examples of SHM?

In analysing real systems that perform SHM to find their time period we always follow the same procedure. We imagine the oscillating mass being displaced from the equilibrium position by displacement x and analyse the forces acting on it as a function of its displacement.

If the system does perform SHM, this resultant force will be a restoring force proportional to x:

- $F = -kx$

where k is a constant of proportionality depending on the parameters of the system. (This equation can be used as an alternative definition of SHM.)

If we now apply (from Newton's second law), we can replace F to write:

- $ma = -kx$
- $a = -\frac{k}{m}x$

Comparing this with the defining equation for SHM

- $a = -\omega^2 x$

we see that

- $\omega^2 = \frac{k}{m}$

This gives us the angular frequency ω, and from this we can obtain frequency f or time period T by

- $f = \frac{\omega}{2\pi}$

or

- $T = \frac{2\pi}{\omega}$

What is the time period of a mass–spring system?

Here, we consider a mass m suspended by a spring of spring constant k. This analysis is complicated a little by the fact that the spring is already stretched when it is in the equilibrium position but, as we shall see, terms that this causes in the equations cancel out, and the analysis ends up looking like the general analysis above.

At equilibrium the tension in the spring is equal and opposite to the weight of the mass

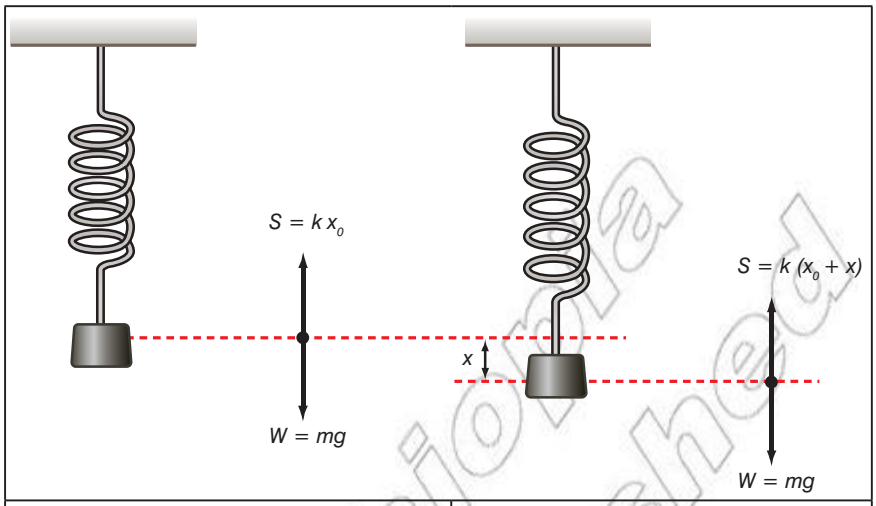
- $S = W$
- $kx_o = mg$

and the resultant force downwards on the mass is zero.

© FDRE, FDR Ethiopia Not to be republished

Think about this...

It is interesting to note the time period of a mass spring system is independent of the gravitational field strength. Take a mass spring system to another planet and the time period of its oscillations will be the same.



(a) Spring and mass at equilibrium with spring stretched so that the tension S in the spring is equal and opposite to the weight of the mass.

(b) Mass displaced x downwards from equilibrium position. Tension in spring now greater than the weight of the mass.

Figure 2.16 *A mass–spring system*

When mass is displaced x downwards, the tension in the spring, and hence the resultant force downwards (in the direction of the displacement), is

- $F = W - S$
- $F = mg - k(x_o + x)$
- $F = mg - kx_o - kx$

But, since $kx_o = mg$, this is

- $F = mg - mg - kx$
- $F = -kx$

Newton's second law tells us that, and so

- $ma = -kx$

and hence

- $a = -\frac{k}{m}x$

Comparing this with the general defining equation for SHM $a = -\omega^2 x$, and recalling that $\omega = \frac{2\pi}{T}$ we have

- $\omega^2 = \frac{k}{m}$
- $\frac{2\pi}{T} = \omega = \sqrt{\frac{k}{m}}$
- $\frac{T}{2\pi} = \sqrt{\frac{m}{k}}$
- $T = 2\pi\sqrt{\frac{m}{k}}$

© MoE Brunei Not to be republished

We have of course assumed that Hooke's law ($S = k \times$ extension) is obeyed. As long as it is, and provided that we can ignore energy losses in the spring and due to air resistance, the mass–spring system performs perfect SHM.

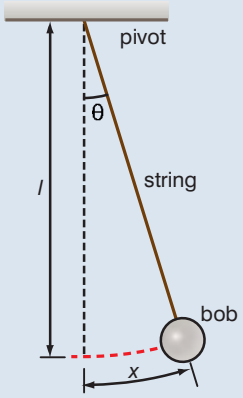
If we make the amplitude of the oscillations too large, however, and we exceed the elastic limit of the spring the above equations are no longer valid and the time period will probably start to become a little longer.

What is the time period of a simple pendulum?

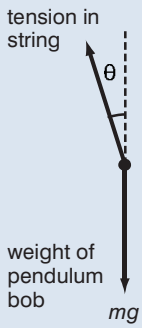
A simple pendulum comprises a single mass m, which we treat as a point mass on a string, length l, (or frictionlessly pivoted rod) whose mass we ignore. This is clearly an approximation and analysis of a simple pendulum is made a little more complicated by the need to make a few more approximations.

To find the time period of a simple pendulum consider the motion of the bob in a circle radius l about the pivot. We analyse the forces acting on the pendulum bob for a displacement x along the circular path that the bob follows, which corresponds to an angular displacement θ, as shown in Figure 2.17. From the definition of angle measurement, in radians

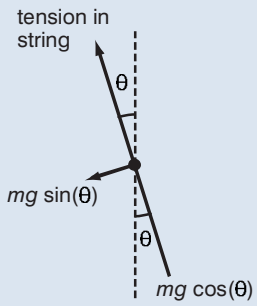
- $\theta = \frac{x}{l}$



(a) Pendulum displaced by angle θ. Note: the pendulum bob follows a circular path



(b) Free body force diagram for pendulum bob.



(c) Free body force diagram with resolved into radial and tangential components relative to circular path of bob

***Figure 2.17** Angular displacement of a pendulum bob*

There are two forces acting on the pendulum bob: the tension in the string acting towards the pivot and the weight on the bob acting vertically downwards. Since the motion of the bob must be at right angles to the string, we know that forces in the direction of the string, radial with respect to the pivot, cannot contribute to the acceleration of the bob along its circular path, at right angles to the string, tangential with respect to the pivot.

If we resolve the weight, mg, of the bob parallel and perpendicular to the string, as shown in Figure 2.17c, the component $mg\cos\theta$

© FDR Ethiopia Not to be republished

parallel to the string makes no contribution to the tangential motion of the bob. The resultant tangential restoring force is the component of the bob's weight perpendicular to the string.

- $F = -mg\sin\theta$

The negative sign tells us that the force is back towards the equilibrium position. The acceleration of the bob is

- $a = \frac{F}{m} = -g\sin\theta$

Now, we have to make another important approximation. For small angles (less than 10°), if we express θ in radians:

- $\theta \approx \sin\theta$

And so, for small angles of swing

- $a = -g\theta = -g\frac{x}{l}$

and so, comparing this to the defining equation for SHM, $a = -\omega^2 x$, we obtain

- $\omega^2 = \frac{g}{l}$
- $\left(\frac{2\pi}{T}\right)^2 = \frac{g}{l}$
- $T = 2\pi\sqrt{\frac{l}{g}}$

It is important to remember what approximations we have made to arrive at the above expression for the time period. We have assumed that the mass of the string can be ignored and that the mass can be treated as a point mass.

For a pendulum where the bob is small compared to the length of the string but has a much greater mass, this is a good approximation. If these assumptions are not valid, then we have a compound pendulum, that requires a different approach, but motion is still SHM. If the angular amplitude of oscillation is not small the approximation that $\theta \approx \sin\theta$ ceases to be valid and the motion, though still periodic, ceases to be SHM. As the amplitude increases, the restoring force for larger displacements will become less required for SHM and the time period will increase.

Big Ben, a famous clock in London, England, has a very large pendulum and the bob has a flat top. Very fine adjustments can be made to its period by adding coins on top of the bob. How does this work?

For a compound pendulum (such as a swinging metre rule) the time period is better expressed using the relationship:

- $T = 2\pi\sqrt{\frac{I}{mgL}}$

where

I = moment of inertia of pendulum

m = mass of pendulum

L = distance from the pivot to the centre of mass of the pendulum.

Discussion activity

Pendulum clocks tend to use quite large masses. Why?



Figure 2.18 *Big Ben in London*

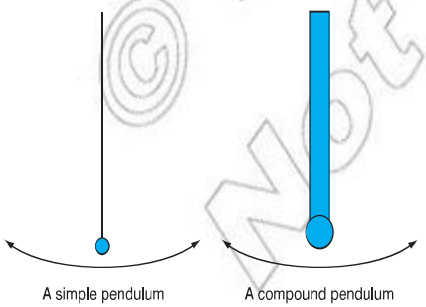


Figure 2.19 *A simple pendulum vs. a compound pendulum*

© NSTA Ethiopia Not to be republished

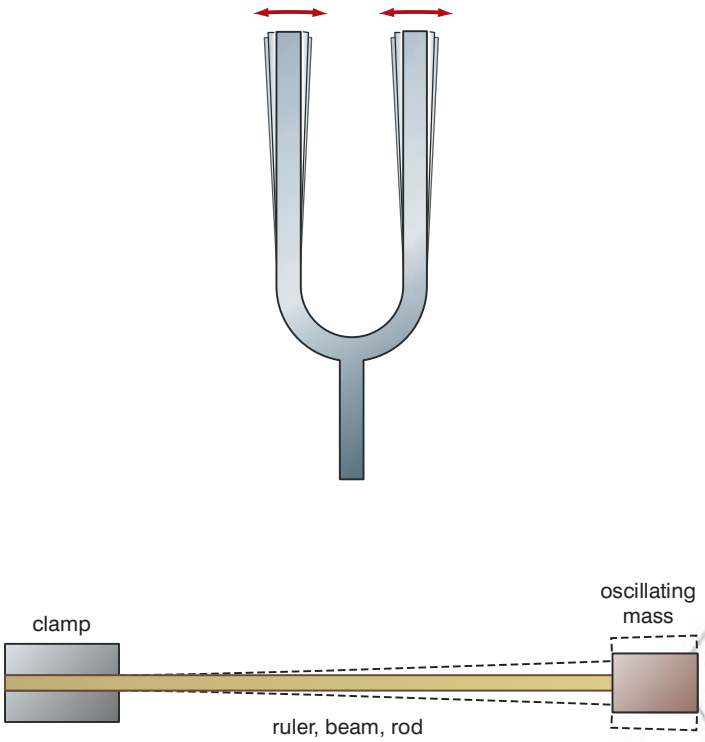


Figure 2.20 *Further examples of systems that perform SHM*

DID YOU KNOW?

When we analyse radial forces acting on the pendulum bob, the tension in the string is only equal to $mg\cos\theta$ when the bob is at the limit of its swing. When the pendulum bob is swinging at velocity v in a circle of radius l, there is a centripetal acceleration towards the pivot and hence a resultant centripetal force:

$$\text{Tension} - mg\cos\theta = \frac{mv^2}{l}$$

Forced oscillations and resonance

A **free oscillation** occurs when an oscillator is displaced from its equilibrium position and released so that it can oscillate freely, with no external forces acting on it.

The oscillator then oscillates at its natural frequency.

Forced or driven oscillations occur when a **periodic driving force** acts on an oscillator. This will make the oscillator oscillate at the frequency of the periodic driving force rather than at its natural frequency. As long as the frequency of the periodic driving force is not the same as the oscillator's natural frequency, the amplitude of the oscillations is usually relatively small.

If the frequency of the periodic driving force is the same as the oscillator's natural frequency, energy is transferred easily into the oscillation and the amplitude of the oscillation becomes large, sometimes very large.

This phenomenon is called **resonance**. Resonance occurs when:

- $f = f_o$

where

f = driving frequency

f_o = natural frequency of the system.

The natural frequency of the oscillator is often referred to as the **resonant frequency**. A plot of driven amplitude against driving frequency peaks at the resonant frequency, as shown overleaf in Figure 2.21.

KEY WORDS

free oscillation *when a body is displaced from its equilibrium position and allowed to oscillate without any external forces acting on it*

periodic driving force *a force of constant frequency acting on an oscillator*

resonance *the tendency of a system to oscillate with larger amplitudes when the frequency of the periodic driving force is the same as the natural frequency of the oscillator*

resonant frequency *the natural frequency of an oscillator*

© MOE FOR Ethi... Not to be republ...

DID YOU KNOW?

In 1940, the Tacoma Narrows Bridge in the USA collapsed within 6 months of being opened after the way the wind flowed over it caused a periodic twisting that ripped it apart. At the time, it was the third longest suspension bridge in the world. This is sometimes described as being classical example of resonance, but this isn't quite true. Simple resonance was already well understood by the bridge designers. The catastrophic vibrations that destroyed the bridge were due to a more complicated phenomenon known as aeroelastic flutter. Lessons learnt from the collapse of the Tacoma Narrows Bridge have affected the designs of suspension bridges ever since.



Figure 2.23 *The Tacoma Narrows Bridge just before its collapse.*

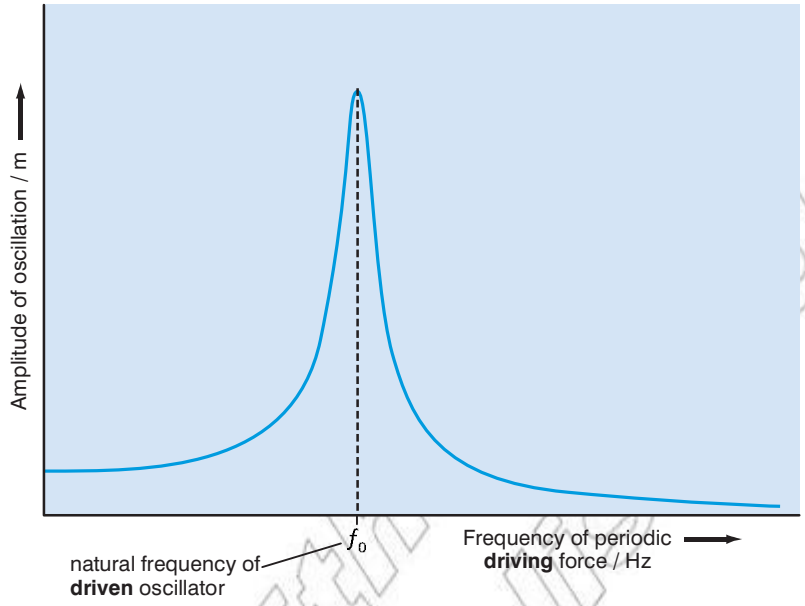


Figure 2.21 *This plot of driven amplitude against driving frequency peaks at the resonant frequency.*

This can be demonstrated using the experimental setup shown in Figure 2.22. The vibrator moves the top of the spring up and down with small amplitude, providing a periodic driving force, with the frequency being set by the signal generator. If the frequency of the signal generator is varied slowly, small oscillations of the mass are observed except at the natural frequency of the mass–spring system, when the oscillations become very large.

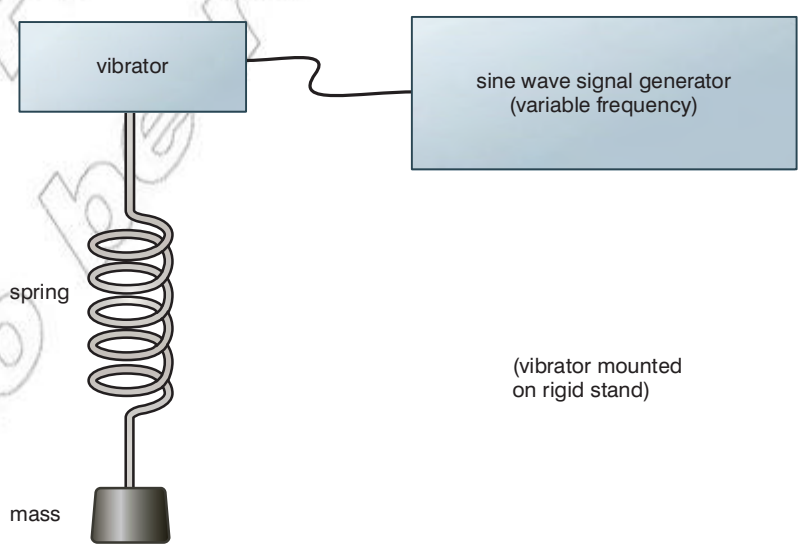


Figure 2.22 *Experiment to demonstrate resonance*

If you've pushed someone on a swing you will be familiar with providing a driving force at the natural frequency of the oscillator. If you stand behind the swing and give a small push just as it reaches the limit of the backward swing you are, by taking your timing from the swing itself, naturally pushing the swing at its natural frequency and, with just gentle pushes, you can quickly build up a large amplitude.

© Free State Department of Education
Not to be republished

If you try pushing the swing at any other frequency, higher or lower, you will find it much harder work to cause even a small amplitude oscillation.

Resonance occurs in many man-made machines and structures. For example, car engines, or unbalanced wheels will create a periodic driving force affecting the whole car, the frequency of this driving force will increase with the car's speed. If this frequency becomes the same as the natural frequency of some part of the car that can oscillate, then that oscillation can become very large, and can be the cause of annoying rattles that occur at specific car speeds.

If resonance does occur, the large amplitude oscillations can cause damage. Bridges can collapse or at least oscillate violently, driven by wind or the regular pace of people walking across them. Troops of marching soldiers often stop marching and walk across bridges out of step to avoid causing this. One way to reduce the effects of resonance and its potentially damaging effects is to design machines so that, if they do oscillate, the natural frequency and the frequency of any periodic driving force are never the same.

A washing machine is a good example of this. The drum that the clothes go into is suspended on springs and can oscillate as a mass–spring system. During the wash cycle the drum revolves slowly, at a rate well below any natural oscillation frequencies, but during the spin cycle the high speed rotation, particularly if the load is unbalanced, could cause a lot of vibration. Most washing machines have a large mass, sometimes made from concrete, strapped to the drum. This extra mass lowers the natural frequency so that it is well below any driving frequency caused by the high speed spin. Sometimes the machine will vibrate violently but very briefly when it starts to spin as the rotation rate passes through the natural frequency.

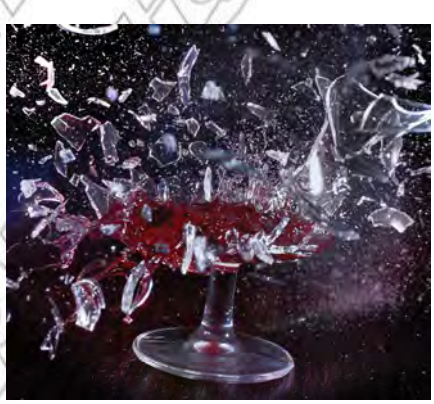


Figure 2.24 *Glasses can be made to shatter if they vibrate at their resonant frequency*

Damping of oscillations

Another way to reduce oscillations is to introduce **damping forces**. Damping forces are resistive, energy dissipating, forces that oppose motion by always being in the opposite direction to the velocity.

Air resistance and friction are typical examples of damping forces and are the reason why pendulums naturally stop swinging and masses on springs stop oscillating.

The damping force is given by:

- $F_d = -bv$

where

b = the damping coefficient and is dependent on the medium providing the damping

v = the velocity of the object through the medium.

This equations shows how the resistive force is directly proportional, but opposite, to the velocity. As a result the amplitude of the oscillation will decay exponentially, as shown overleaf in Figure 2.25 (a). Note that the period of the oscillation does not

KEY WORDS

damping forces *resistive forces that oppose the motion of an oscillator by acting in the opposite direction to its velocity*

© Mobile Ethiopia Not to be republished

Discussion activity

Once a suspension bridge has been built it is very difficult to change its natural frequency of oscillation. Why?

change as the amplitude gets smaller. Heavier damping causes a more rapid decay of amplitude as shown in Figure 2.25(b).

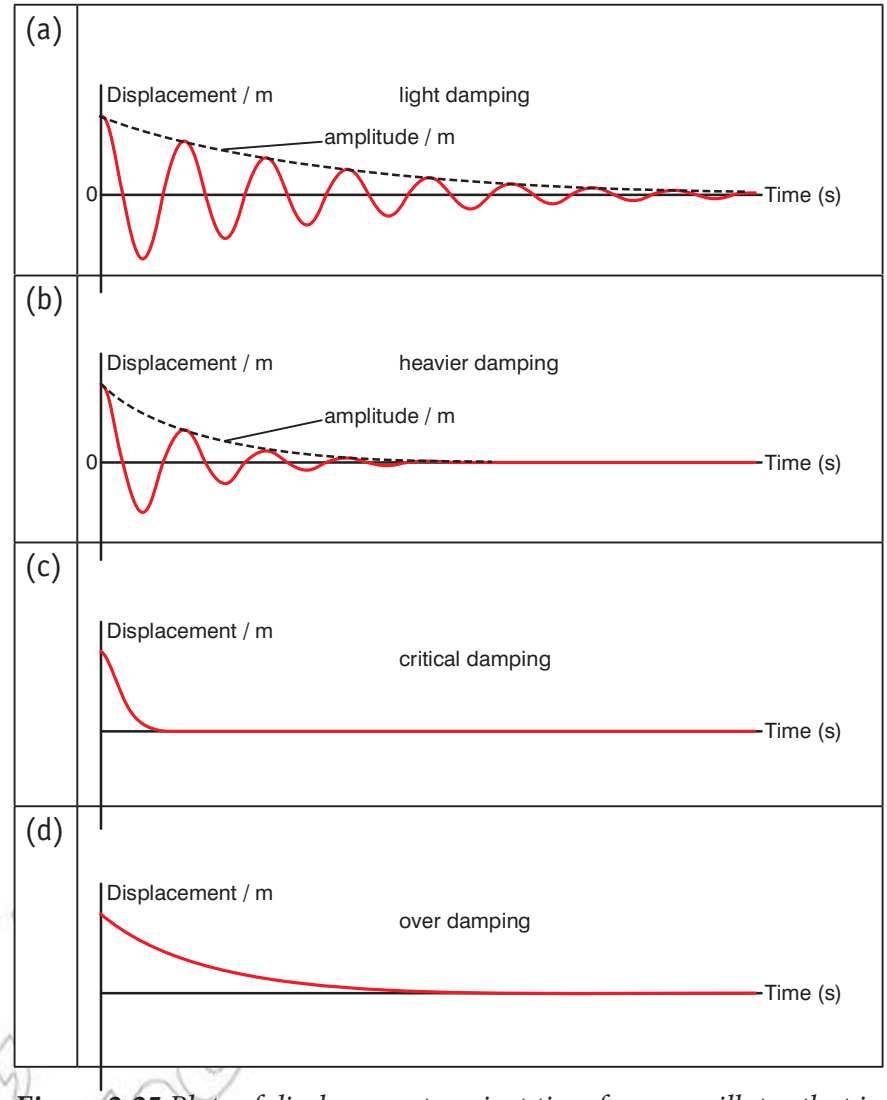


Figure 2.25 *Plots of displacement against time for an oscillator that is displaced and then released, for different amounts of damping.*

KEY WORDS

overdamping *damping that prevents oscillation entirely and only allows the oscillator to return slowly to its equilibrium position*

An example of deliberate damping can be found in a car suspension system. A piston inside cylinder, as shown in Figure 2.26, containing viscous oil can move but the faster it moves the greater the resistance to movement. If such damping is very heavy it can prevent oscillation altogether, so that if the 'oscillator' is displaced it can only return very slowly to the equilibrium position. This is known as **overdamping** and is shown in Figure 2.25(d).

Activity 2.4: Damping

Identify the type of damping in the following cases and justify your answer.

a) Pendulum in air

b) Pendulum in water

c) Pendulum in thick treacle

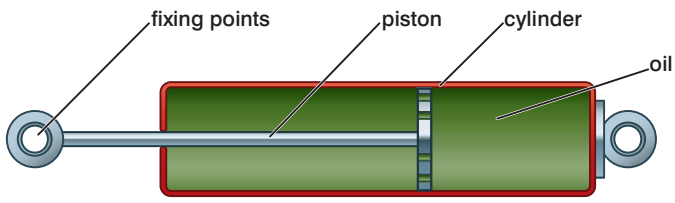


Figure 2.26 *A simple viscous damper. The piston can move inside the cylinder but the faster it moves the greater the resistance to movement.*

Damping in a car suspension is not normally so heavy, as this would produce a very 'hard' and uncomfortable ride for the passengers. The damping shown in Figure 2.25(b), on the other hand, would provide a very bouncy ride; this would be called **underdamping**. The damping in a car suspension is always a compromise somewhere near to the critical damping shown in Figure 2.25(c). **Critical damping** is the amount of damping that leads to the oscillator settling back to a stationary state at the equilibrium position in the shortest possible time.

Damping reduces the effects of resonance. As the periodic driving force transfers energy into the oscillator the damping mechanism dissipates the energy. The resonance peak in the graph of driven amplitude against driving frequency becomes lower and relatively wider, as shown in Figure 2.27. It can also be seen that damping also causes a very small reduction in the natural frequency of the oscillator.

KEY WORDS

underdamping *damping that allows the oscillator to move back and forth through its equilibrium position before returning to rest*

critical damping *the amount of damping that allows the oscillator to return to its equilibrium position in the shortest possible time*

conservation of energy *the total amount of energy in an isolated system remains constant over time*

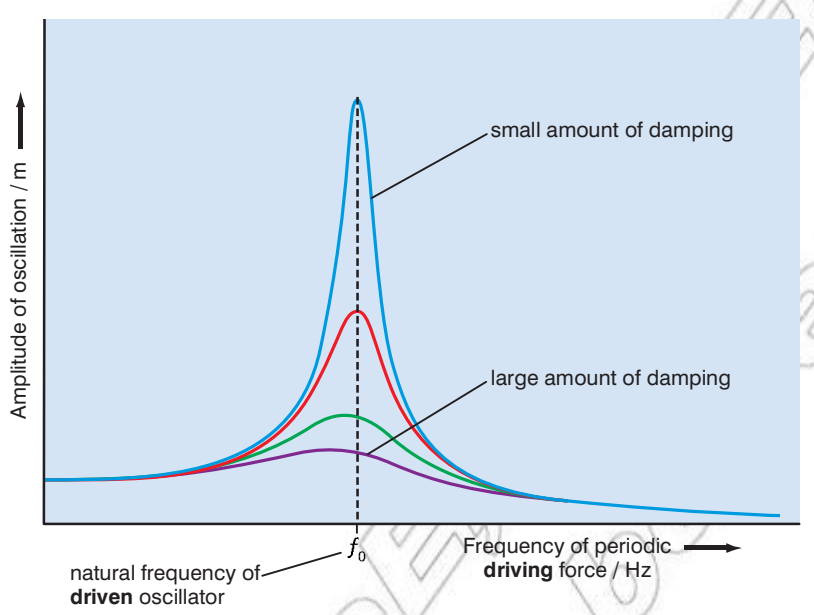


Figure 2.27 *Driven amplitude against driving frequency for forced oscillations of an oscillator with different amounts of damping.*

Energy in SHM

Any oscillator performing SHM has energy and the law of **conservation of energy** tells us that, in the absence of any external forces or damping, that energy must be constant even if it may be changing in form.

When the oscillator is passing through the equilibrium position, when , the resultant force acting on it is zero and it has no potential energy but it is moving at the maximum velocity and has kinetic energy. When the oscillator is at the limit of oscillation, when $x = A$, and the velocity is temporarily zero, the kinetic energy must be zero, but the force acting on the oscillator is at a maximum and the oscillator's energy is all stored as potential energy.

© Maskew Miller Longman (Pty) Ltd
Not to be republished

We know that the kinetic energy at any time is given by

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

To obtain an expression for the potential energy (*PE*) of the oscillator for any displacement x we need to calculate how much work has been done against the restoring force to move it to that displacement. We can do using the relationship work equals force time distance, but we have to take into account that the restoring force is not one constant value but increases with x.

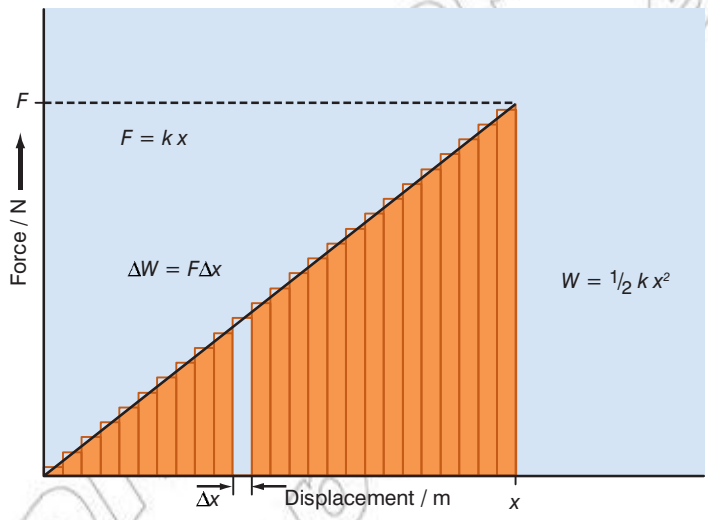


Figure 2.28 *Calculation of PE. The work done against the restoring force in moving the oscillator from the equilibrium position to displacement x is the area under the graph of force against displacement.*

Figure 2.28 shows how we can calculate the work done. For any small increase Δx in x the work done is $\Delta W = F\Delta x$, where $F = kx$ is the force given by . We can see that this contribution to the total work done is a portion of the total area under the graph of force against displacement and that the total work done is the total area under the graph, and hence

- $PE = \frac{1}{2}kx^2$

The total energy at any time is the sum

- $Total\ energy = PE + E_k = \frac{1}{2}kx^2 + \frac{1}{2}mv^2$

This total energy is equal to the kinetic energy when $x = 0$ and $v = v_0$, or to the potential energy when $x = A$, i.e.

- $Total\ energy = \frac{1}{2}mv_0^2 = \frac{1}{2}kA^2$
- $Total\ energy = \frac{1}{2}m\omega^2A^2$

Note that the total energy of an oscillator is proportional to the amplitude squared.

© MoE/EDI Not to be reproduced

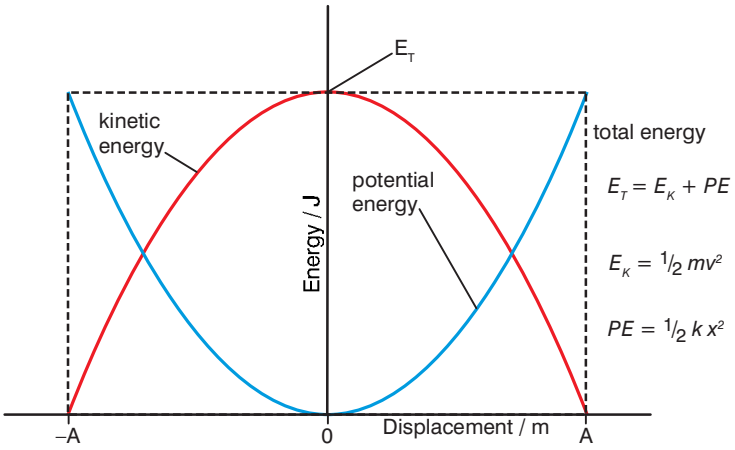


Figure 2.29 *Variation of kinetic energy and potential energy of an oscillator with displacement, showing that the sum remains constant*

This variation of E_k and PE with displacement is shown in Figure 2.29.We can show that these expressions for energy are consistent with our earlier expression for the value of velocity v for any given displacement x:

- $v = \omega\sqrt{A^2 - x^2}$ Starting with our equation for velocity
- $v^2 = \omega^2(A^2 - x^2)$ Making v^2 the subject
- $v^2 = \omega^2 A^2 - \omega^2 x^2$ Multiplying out the brackets

As this is a mass–spring system we can substitute in k/m for ω^2, giving

- $v^2 = \frac{k}{m}A^2 - \frac{k}{m}x^2$

Substituting v^2 into our E_k equations gives:

- $\frac{1}{2}mv^2 = \frac{1}{2}kA^2 - \frac{1}{2}kx^2$

Therefore

- $E_k = total\ energy - PE$

Worked example 2.3

A block of mass 2.2 kg is attached to a spring with a spring constant of 40 N/m. It is pulled down a distance of 30 cm. Find the blocks kinetic energy as it passes through the equilibrium and determine its velocity at this point.

- $\frac{1}{2}mv^2 = \frac{1}{2}kA^2 - \frac{1}{2}kx^2$

Using the relationship above, but is this case as the block is passing through its equilibrium position x = 0 so the relationship simplifies to:

- $\frac{1}{2}mv^2 = \frac{1}{2}kA^2$

Substitute in known values and solve, giving:

- $E_k = ½kA^2 = ½ \times 40 \times 0.3^2 = 1.8$ J

FDRE, Ethiopia
be republished

The velocity can then easily be determined using the equation for kinetic energy:

- $E_k = \frac{1}{2}mv^2$
- $v = \sqrt{2E_K/m}$ *Rearrange to make v the subject*
- $v = \sqrt{((2 \times 1.8) / 2.2)}$ *Substitute known values*
- $v = 1.3$ m/s *Solve and give the unit*

Summary

In this section you have learnt that:

- Simple harmonic motion (SHM) is a periodic oscillation of an object about an equilibrium position such that its acceleration is always directly proportional in size but opposite in direction to its displacement. The defining equation is
  $a = -\omega^2 x$
  where
  $\omega = 2\pi f = \frac{2\pi}{T}$.
- For an oscillator performing SHM:

  time period does not depend on amplitude

  $x = A\sin(\omega t)$

  $v = \omega A\cos(\omega t)$

  $a = -\omega^2 A\sin(\omega t)$

  $v = \omega\sqrt{A^2 - x^2}$
- For a mass–spring system $T = 2\pi\sqrt{\frac{m}{k}}$.
- For a pendulum $T = 2\pi\sqrt{\frac{l}{g}}$.
- An oscillator will oscillate at its natural frequency if displaced and allowed to oscillate freely without external forces and damping.
- Forced oscillations occur if an oscillator is driven by a periodic driving force.
- Resonance occurs when an oscillator is driven at its natural frequency
- Damping forces are opposite in direction to velocity and dissipate energy, causing an exponential decay in the amplitude of free oscillations.
- Damping reduces the amplitude of driven oscillations, reducing the effects of resonance
- Total energy = $PE + E_k = \frac{1}{2}kx^2 + \frac{1}{2}mv^2$
- The total energy of an oscillator is proportional to amplitude squared.

© MoE Not for

Review questions

1. An object moving with simple harmonic motion has an amplitude of 3 cm and a frequency of 30 Hz. Calculate:
   a) the time period of the oscillation,
   b) the acceleration in the centre and at the maximum displacement of an oscillation, and
   c) the velocity in the centre and at the maximum displacement of an oscillation.
2. How long does a pendulum need to be to have a time period of 1 second? Explain all the approximations and assumptions you make to carry out this calculation.
3. A mass of 500 g is suspended on a vertical spring of spring constant $k = 10$ N/m.
   a) Calculate the frequency at which the mass will oscillate if displaced downwards a small distance and released.
   b) A periodic driving force of variable frequency f is applied to the top of the spring. Sketch and explain a graph of amplitude against frequency f for the oscillation of the mass.
   c) On the same axes, sketch a graph of amplitude against frequency f for the oscillation of the mass if a relatively large piece of cardboard is taped horizontally to the mass.
   d) On the same axes, sketch a graph of amplitude against frequency f for the oscillation of the mass if its size is increased to 1 kg.
4. A simple pendulum has a length of 1.2 m and the bob has a mass of 800 g. The pendulum swings with an amplitude of 14 cm. Calculate:
   a) the velocity of the pendulum bob at the centre of its swing
   b) the kinetic energy of the pendulum bob at the centre of its swing
   c) the kinetic energy and the potential energy of the pendulum bob when it is 8 cm from the centre of its swing.
5. Describe the key features of the different forms of damping the general effect of damping on resonance.

© MoE-FDRE Ethiopia Not to be republished

2.2 Wave motion

By the end of this section you should be able to:

- Describe the characteristics of a mechanical wave and identify that the speed of the wave depends on the nature of medium.
- Use the equation $v = \sqrt{T/\mu}$ to solve related problems.
- Describe the characteristics of a travelling wave and derive the standard equation $y = A\cos(\omega t + \phi)$
- Define the terms phase, phase speed and phase constant for a travelling wave.
- Explain and graphically illustrate the principle of superposition, and identify examples of constructive and destructive interference.
- Identify the properties of standing waves and for both mechanical and sound waves, explain the conditions for standing waves to occur. Including definitions of the terms node and antinode.
- Derive the standing wave equations.
- Calculate the frequency of the harmonics along a strong, a open pipe and a pipe closed at one end.
- Explain the modes of vibration of strings and solve problems involving vibrating strings.
- Explain the way air columns vibrate and solve problems involving vibrating air columns.
- Analyse, in quantitative terms, the conditions needed for resonance in air columns, and explain how resonance is used in a variety of situations.
- Identify musical instruments using air columns, and explain how different notes are produced.



Figure 2.30 *Waves on water*

KEY WORDS

direction of propagation *the direction in which energy is transferred along a travelling wave*

mechanical wave *a wave that involves the oscillations of particles of a physical medium*

What is a travelling wave?

Electromagnetic and sound waves are particularly important to us, but waves on water are a little easier to observe. If we drop a pebble into a pond we see small waves or ripples radiating outwards.

If dip a stick into the middle of the pond and move it up and down with SHM the motion becomes continuous. These waves are spreading out in two dimensions and sound and electromagnetic waves spread out in three dimensions, but for the moment, we will simply consider waves travelling in one dimension. We will later consider stationary, or standing, waves, but first we need to understand a little about travelling waves.

A travelling wave transfers energy, and sometimes information, from one place to another, in what is called the **direction of**

propagation. An oscillation at the source of energy causes an oscillation to travel through space. For electromagnetic waves this oscillation is of electric and magnetic fields and does not need a medium. In a **mechanical wave** that involves the oscillations of particles of a physical medium, as the particles pass on energy, they undergo temporary displacements but no permanent change in the position. For example, when ripples travel across a pond the water molecules oscillate vertically but do not move in the direction of the wave.

Transverse and longitudinal waves?

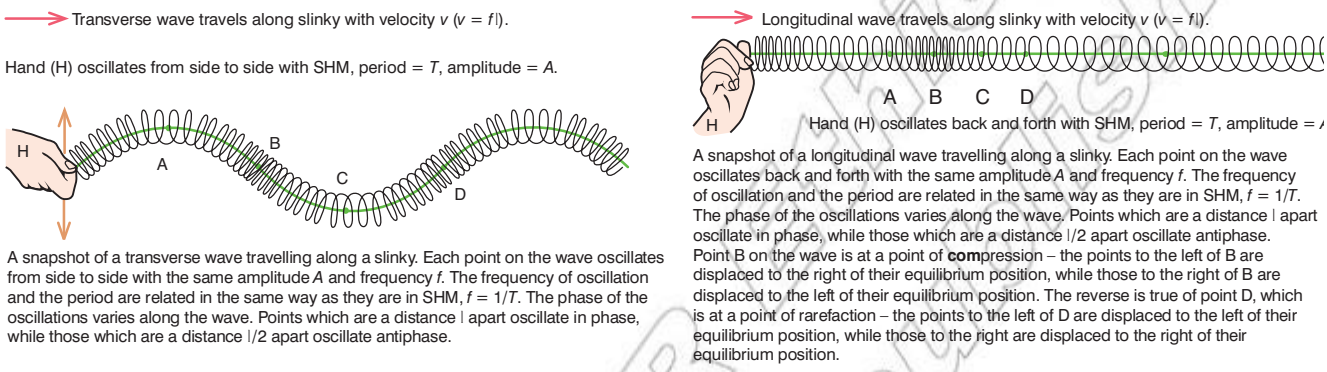


Figure 2.31 *Waves along a slinky*

We can demonstrate a wave travelling along a stretched slinky, as shown in Figure 2.31. We can create two distinctly different types of travelling wave.

A **transverse wave** is one in which the oscillation, the temporary displacement of mass or field strength, is at right angles to the direction of propagation. Electromagnetic waves and waves travelling along a string or rope are examples of transverse waves. Waves on water can appear to be transverse if the amplitude is small, but in reality they are more complicated and involve the water moving in circles.

KEY WORDS

transverse wave *wave where the oscillations are perpendicular to the direction of wave motion*

longitudinal wave *wave where the oscillations are parallel to the direction of wave motion*

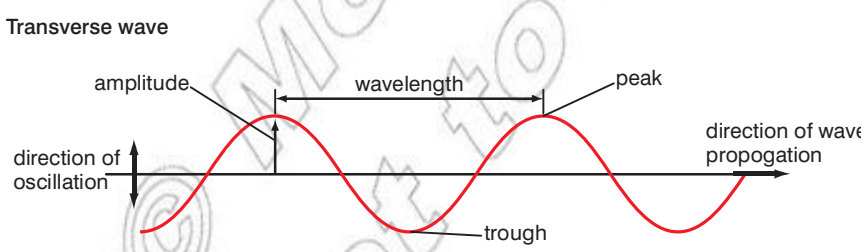


Figure 2.32 *A transverse wave*

A **longitudinal wave** is one in which the oscillation, the temporary displacement of mass, is backwards and forwards along the path of wave propagation/net energy transfer. Sound is a longitudinal wave.

© Motion EDR Namibia Not to be reproduced

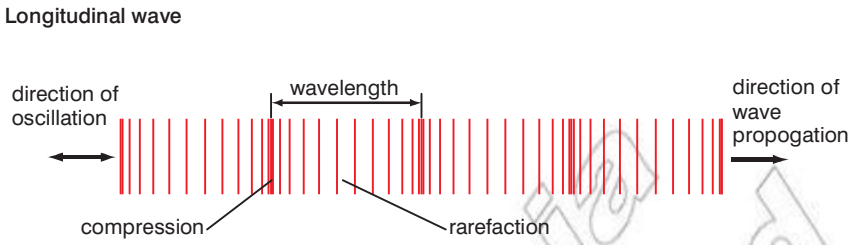


Figure 2.33 *A longitudinal wave*

A wave is carried by a chain of oscillators, each passing on its energy to the next oscillator. Hence, just as the energy of a single oscillator is proportional to the square of the amplitude of the oscillation, the power or rate of energy transfer of a wave is proportional to amplitude squared.

Wave speed

The frequency of a wave can be defined in two equivalent ways. It is the frequency of the individual oscillators that pass the energy along, the number of times particles go up and down or backwards and forwards per second.

It is also the number of complete waves, the number of wavelengths that pass any given point per second. If the wavelength is λ, and f wavelength pass a point per second, then the speed of the wave must be given by the wave equation:

- $v = f\lambda$

Wave speed through different media

The speed of any travelling wave depends on the media it is travelling through (more on the speed of sound in chapter 2.3).

For a mechanical wave travelling along a string the speed of the wave depends on the **tension** of the string and the **mass per unit length** (sometimes called **linear density**).

- $v = \sqrt{\frac{T}{\mu}}$

where

μ = mass per unit length given by $\mu = \frac{m}{l}$ in kg/m

T = tension in the string in N.

The formula given above shows us that the 'tighter' the string the faster the waves will travel down its length. Additionally the 'lighter' the string, (the smaller its mass/length ratio), the faster the waves will travel down its length.

The **phase speed** of a wave is the rate at which the phase of the wave travels through space. Any given phase of the wave (for example, the crest or the trough) will appear to travel at the phase velocity. The phase velocity is given in terms of the wavelength λ (lambda) and period T as

- $v_{phase} = \lambda / T$

Worked example 2.4

If a sound wave with a frequency of 500 Hz passes through a liquid at a speed of 1500 m/s, then its wavelength must be

- $\lambda = \frac{v}{f}$
- $\lambda = \frac{1500}{300} = 3$ m

KEY WORDS

tension *a measure of the force tending to stretch a string*

mass per unit length *a measure of the distribution of the mass of a string along its length*

linear density *a measure of mass per unit length*

phase speed *the rate at which the phase of the wave travels through spa*

© MOE Free Online Not to be resold Finished

How do we describe a travelling wave mathematically?

An oscillation at the source causes a travelling wave, which causes oscillations along its path. A mathematical description of the wave must give an expression for the temporary displacement Y at any distance x along the path of the wave at any time t.

If a wave is sinusoidal, then a snapshot of the wave, i.e. a side view at an instant in time, looks as shown below in Figure 2.34(a). If the wave now travels on from the position shown it causes particles to oscillate with SHM at points A and B but, because the wave has to travel a quarter of a wavelength further to reach point B, the oscillation at point B always lags behind that at point A by quarter of a cycle, $\pi/2$ radians, as shown in Figures 2.34 (b) and (c).

If points A and B were half a wavelength apart than the oscillation at B would lag half a cycle, π radians, behind that at A. If A and B were a whole wavelength apart that the oscillation at B would be a whole cycle behind that at A and therefore be effectively back in step with it.

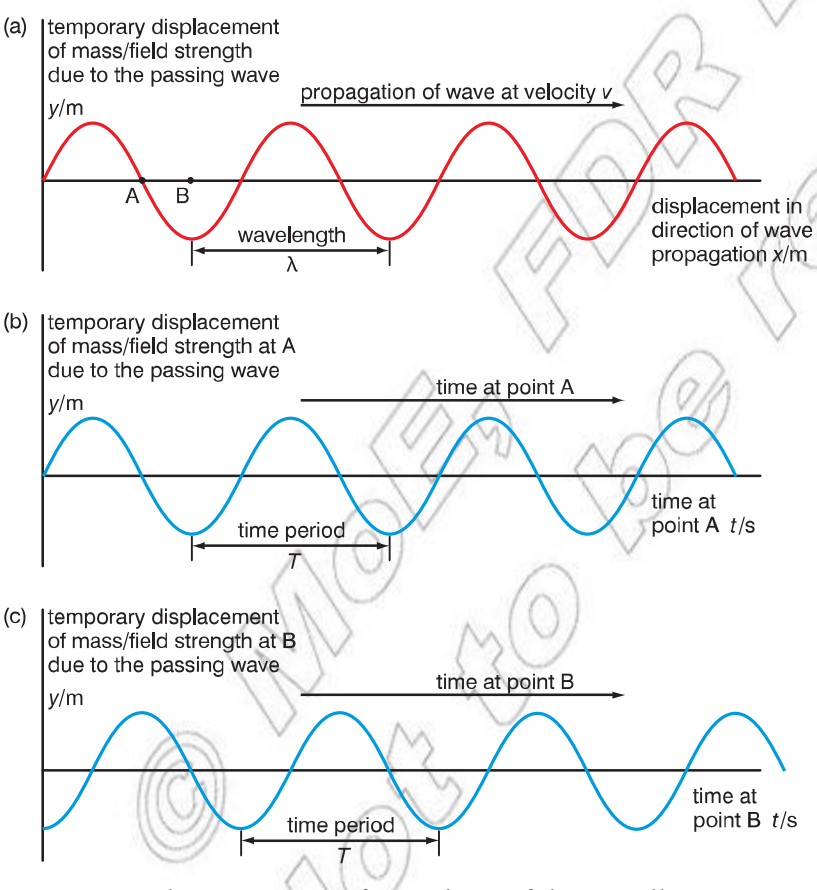


Figure 2.34 *A side view of a transverse wave at a single instant in time. As the wave is sinusoidal, the wave causes particles to oscillate with SHM at points A and B. Over time there is a phase shift, this does not change the shape of the move but moves it back and forth along the x-axis.*

Figure 2.34 shows a series of snapshots of the travelling wave at successive instants in time, quarter of a cycle apart. At time $t = 0$, the wave is a sine wave described by the equation

- $Y = A\sin\left(2\pi\dfrac{x}{\lambda}\right)$

When $t = \dfrac{T}{4}$, the wave has moved a quarter of a wavelength to the right and is described by the equation

© MoE FDR Ethiopia Not to be republished

- $Y = A\sin\left(2\pi\frac{x}{\lambda} - \frac{\pi}{2}\right)$

We can confirm that this is the correct expression by checking the values that this gives.

When $x = 0$ $\sin\left(2\pi\frac{x}{\lambda} - \frac{\pi}{2}\right) = \sin\left(-\frac{\pi}{2}\right) = -1$

When $x = \frac{\lambda}{4}$, $\sin\left(2\pi\frac{x}{\lambda} - \frac{\pi}{2}\right) = \sin\left(2\pi\frac{1}{4} - \frac{\pi}{2}\right) = \sin(0) = 0$

and so on.

When $t = \frac{T}{2}$, the wave has moved a half of a wavelength to the right and is described by the equation

- $Y = A\sin\left(2\pi\frac{x}{\lambda} - \pi\right)$

In general, after time t:

- $Y = A\sin\left(2\pi\frac{x}{\lambda} - 2\pi\frac{t}{T}\right)$

or

- $Y = A\sin\left(2\pi\frac{x}{\lambda} - 2\pi ft\right)$

This is a very useful description. By substituting in a value of x for the position of a point along the wave's path, we can obtain an expression for the oscillation at that point, or by substituting in a value for t at a particular instant we can obtain an expression describing the shape of the wave at that instant.

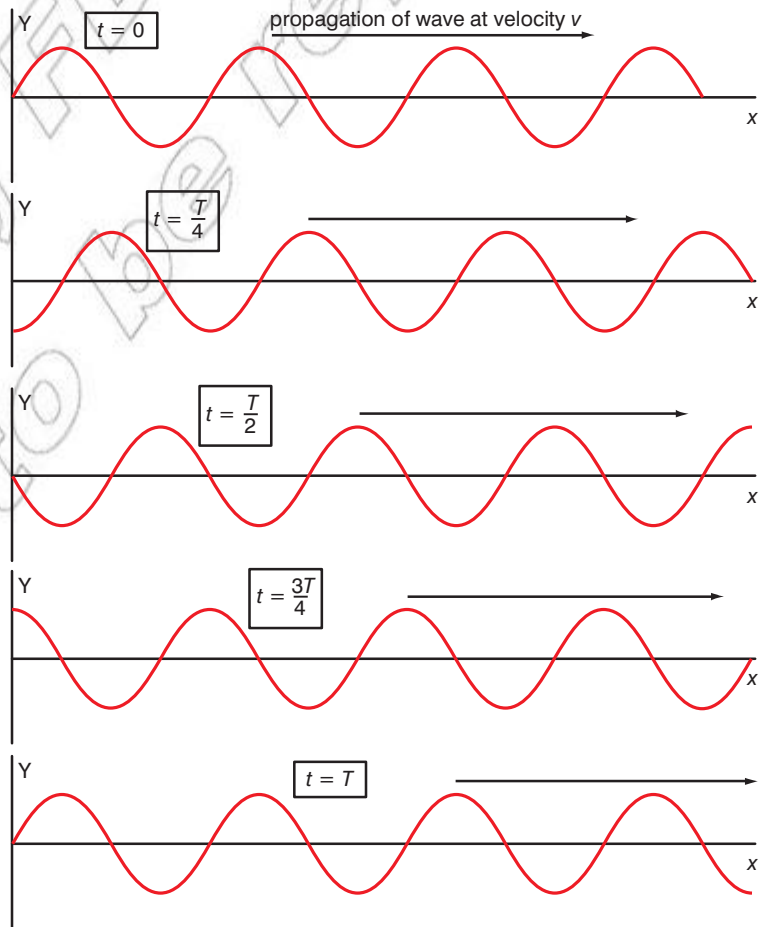


Figure 2.35 *A travelling wave shown at a series of successive times*

© MoE: FDRE, Ethiopia
Not to be republished

Just like SHM it does not really matter whether a sine or cosine function is used to describe the travelling wave. A cosine version of the travelling wave equation may be seen below:

- $Y = A\cos(\omega t + \phi)$

In this equation ϕ is the phase constant of the wave, this effectively moves the waves back and forth along the x-axis.

Principle of superposition

If two or more waves pass through a single point then the resultant instantaneous displacement at that point is the sum of the displacements that would be created separately by each wave, taking signs into account. The waves pass on through the point and each other and continue on unaffected. This works for two waves passing through each other at any angle, as shown in Figure 2.36, and for waves passing through each other in opposite directions along a string, as shown in Figure 2.37.

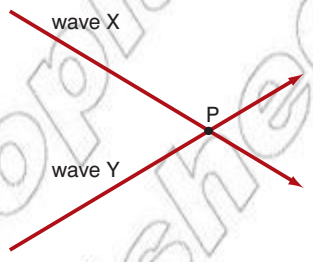
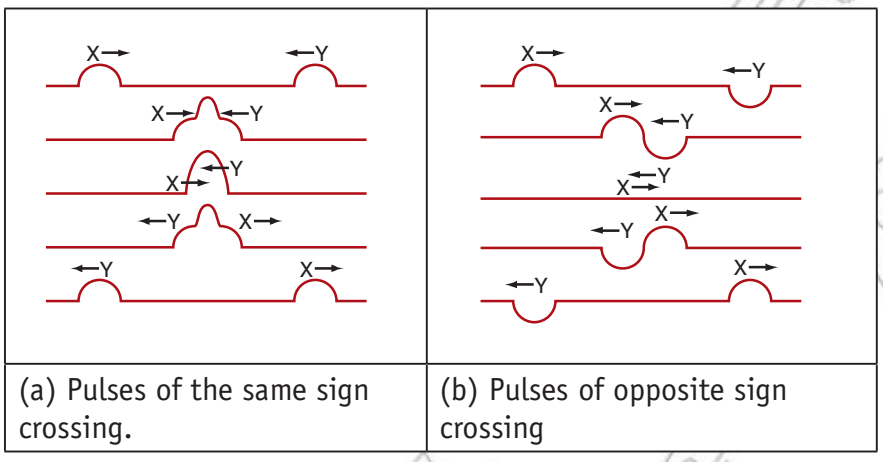


Figure 2.36 *Two waves, X and Y, pass through each other unaffected, but the instantaneous displacement at point P is the sum of the of the displacements that would be created separately by each wave.*



(a) Pulses of the same sign crossing.	(b) Pulses of opposite sign crossing

Figure 2.37 *Pulses passing through each other going in opposite directions along a string and demonstrating superposition as they cross*

Constructive and destructive interference

Interference is another word for superposition. When two waves arrive at the same point they interfere with each other: the instantaneous displacement at that point is the sum of the displacements that would be created separately by each wave. If we consider just two waves, the result is an oscillation whose amplitude depends on the relative **phase** of the two waves.

The **phase difference** between two oscillations is an angular measurement of the difference in their timing, best understood by thinking about the circular motion link to SHM shown in Figure 2.11. Since a whole cycle involves an angular change of 2π radians, a half cycle difference is a phase difference of π, and a quarter cycle difference is a phase difference of $\pi/2$.

If the two waves are causing independent oscillations that go up and down, and pass through the equilibrium point at the same times, as shown overleaf in Figure 2.38, they are said to be in phase with each other.

KEY WORDS

phase *a measurement of the position of a point on a wave after a particular time. Two sine waves are said to be in phase when corresponding points of each reach maximum or minimum displacements at the same time.*

phase difference *the angular difference in timing between 2 waves*

© Kwa Ethiopia Not to be published

KEY WORDS

antiphase *where two sine waves are performing the opposite motion to each other. The phase difference between them is 180 degrees.*

constructive interference *the production of large oscillations by the superposition of two waves that are in phase with each other*

destructive interference *the cancelling out of oscillations caused by the superposition of two waves that are in antiphase*

If they are always performing completely the opposite motion to each other, as shown in Figure 2.38, they are said to be completely out of phase, 180° or π radians out of phase with each other, or in **antiphase**.

If two oscillations are in phase with each other we get **constructive interference** giving a large amplitude oscillation.

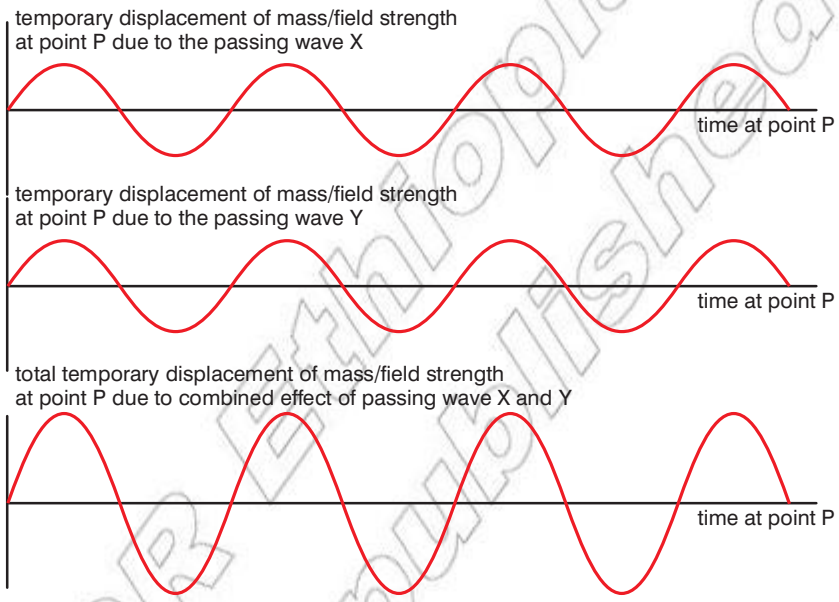


Figure 2.38 *Constructive interference: two oscillations in phase with each other combine to produce a larger oscillation at the same frequency.*

Think about this...

Waves at sea are caused by the action of wind on the surface of the water. In big oceans on a calm day you see big waves that have travelled thousands of kilometres from a storm centre thousands of kilometres away. Sometimes these waves are nice and regular. Sometimes you can experience a very irregular pattern of waves getting stronger and weaker and very difficult to predict. Why is this?

If the two oscillations are in antiphase we get **destructive interference** or cancellation leading to a small or zero resultant oscillation. Note that we only get complete cancellation if the two oscillations are in perfect antiphase and have the same amplitude.

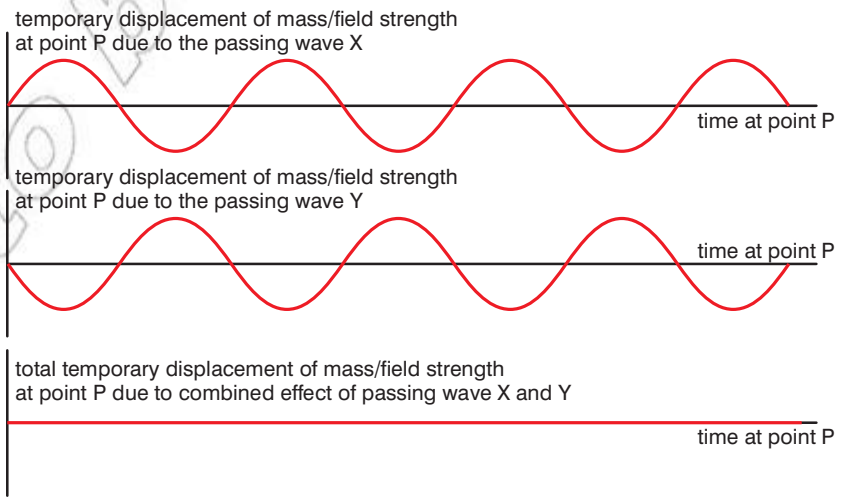


Figure 2.39 *Destructive interference: two oscillations in antiphase cancel each other to produce a small or zero resultant oscillation.*

NOT to be republished © Ethiopia

Reflections of waves

When a travelling wave reaches a sudden change in medium it will be at least partially reflected. In some circumstances this reflection can be total. When a wave travelling along a string reaches the end, it will be completely reflected if the end of the string is either firmly clamped so that it cannot move at all or if the end is completely free to move. If the end of the string is connected to a second string of different mass per unit length, some energy will be reflected and some will be transmitted on along the next string.

If the end of the string is fixed so that it cannot oscillate at all, then the sum of the incident and reflected wave must be always zero and so the reflected wave must be in antiphase. The reflection includes a phase shift of π radians. If the end of the string is completely free to move, we still get 100% reflection, but with no phase shift. The same rules apply to sound waves in narrow tubes, as in musical instruments. Where the end of the tube is closed there is no net oscillation and the sound wave is reflected with a phase shift of π radians; where the end of the tube is open there is a large net oscillation and the sound wave is reflected with no phase shift.

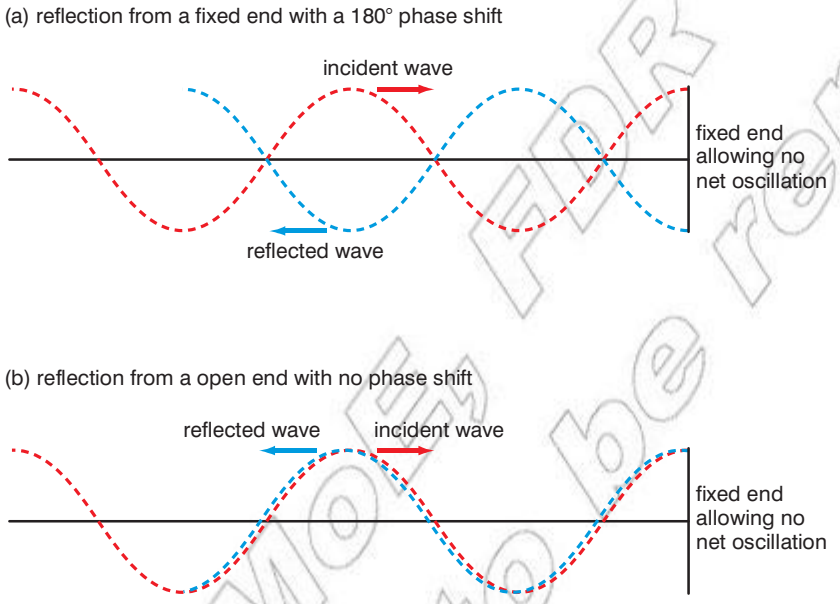


Figure 2.40 *Reflections at the end of a string, or air column*

Standing waves on strings

Musical instruments use standing, or stationary, waves, either on strings or in air columns, to generate sound at different frequencies. If a string is fixed at both ends and it is plucked, or has a bow drawn across it, then waves will travel away along the string to be reflected at the ends. This produces waves travelling in opposite directions along the string and they will interfere with each other. The waves that travel away from the initial point of plucking will be at a wide range of different frequencies, and for most of these frequencies the reflections will never interfere constructively with each other and they will disappear. At some specific wavelengths

© MoE FDR Ethiopia Not to be republished

however, depending on the length of the string, the waves travelling in opposite directions will interfere to produce a standing wave, as shown in Figure 2.41. If we can observe the oscillation of the string, which can be done using high speed photography or using a stroboscope, we do not see travelling wave but a wave shape that stays in one place, a stationary, or standing, wave.

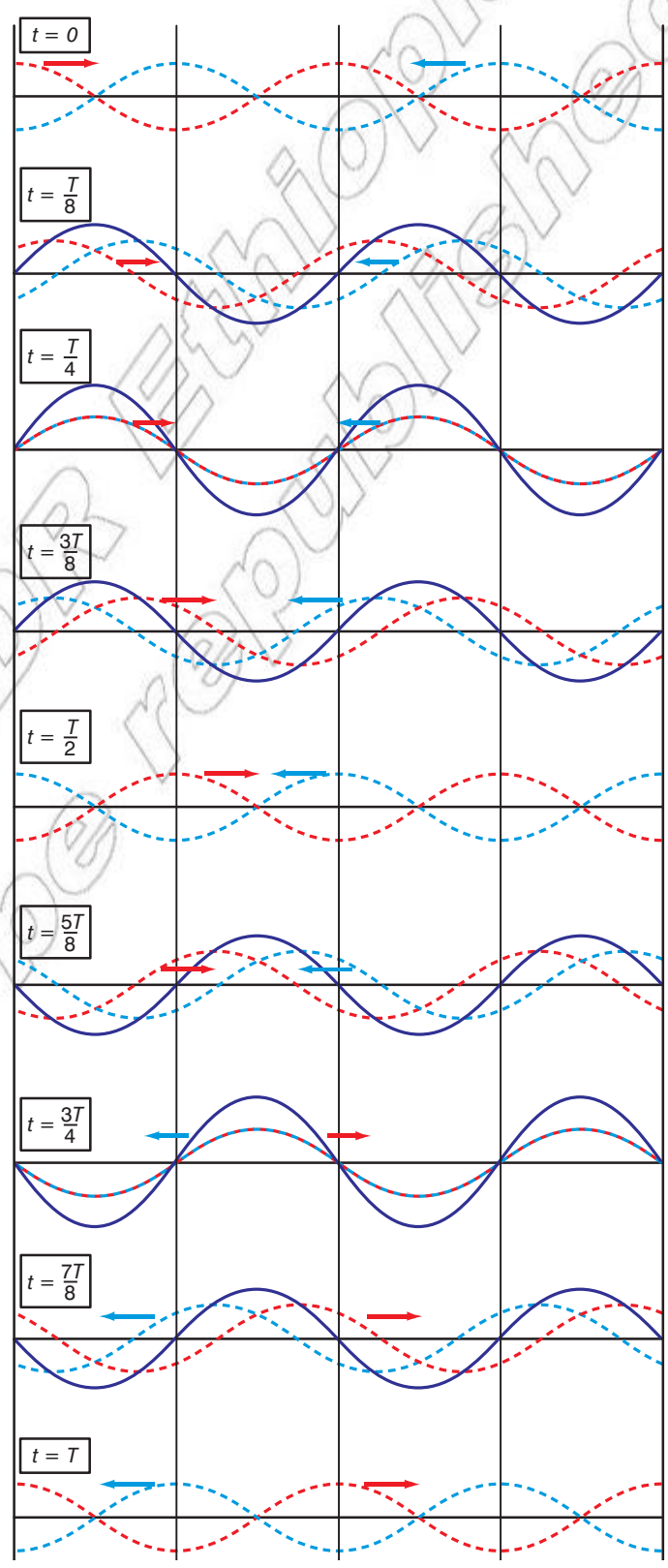


Figure 2.41 *Formation of a stationary wave by superposition of two waves of equal amplitudes and wavelength travelling in opposite directions .The red and blue waves are the two travelling waves, which we do not see in reality. The black line is the standing wave that we do see.*

© MoE, FDR Ethiopia
Not to be republished

At points where the wave travelling to the right and the wave travelling to the left are always in antiphase, as at the fixed ends, superposition produces no net oscillation. These points are called **nodes**. There are positions however where the two waves are always in phase with each other and here superposition produces large oscillations.

The points where the two waves are always perfectly in phase and the net oscillations largest are called **antinodes**, as shown in Figure 2.42. All points of the string between any two adjacent nodes, half a wavelength, oscillate in phase with each other.

KEY WORDS

nodes *the points where two superimposed waves are in antiphase and there is no net oscillation*

antinodes *the points where two superimposed waves are in phase and the net oscillations are largest*

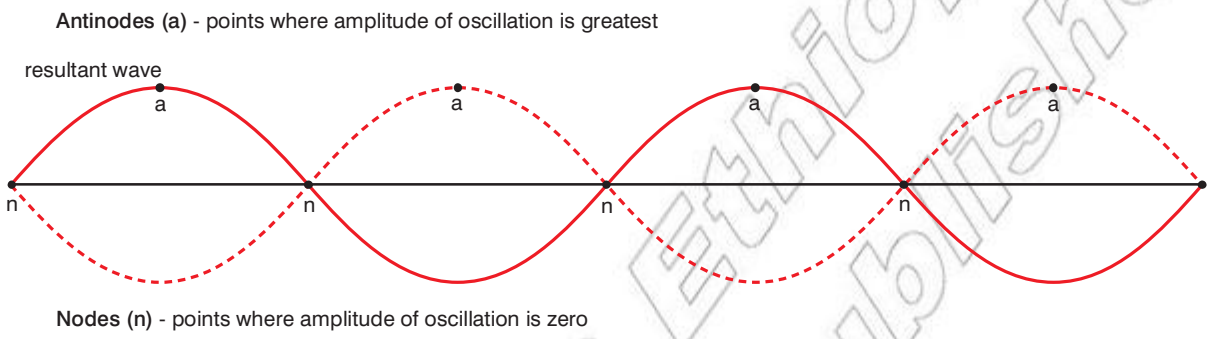
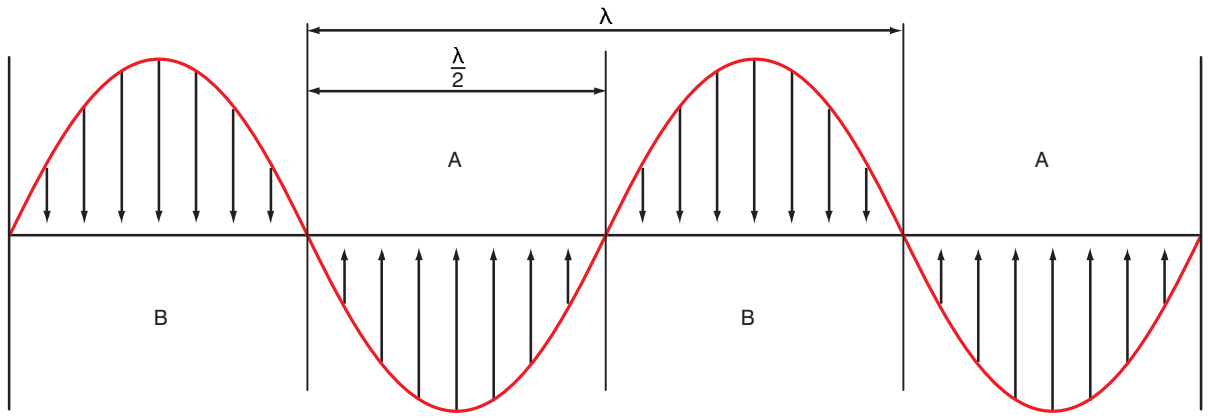


Figure 2.42 *Nodes and antinodes on a stationary wave. The distance between two successive nodes and antinodes is half a wavelength.*

Adjacent half wave length sections are in antiphase with each other, as shown in Figure 2.43.



All particles in the string in sections A oscillate in phase with each other – those in the centre with the biggest amplitude	All particles in the string in sections B oscillate in phase with each other – those in the centre with the biggest amplitude
Particles in section A oscillate in anti-phase with particles in section B, i.e. their oscillations have a phase difference of π or 180°.	

Figure 2.43 *Oscillations in different sections of a stationary wave*

The location of nodes and antinodes are very important.

- Nodes occur when the distance along the string is equal to $n\lambda / 2$
- Antinodes occur when the distance along the string is equal to $(n+½)\,\lambda / 2$.

Where n = an integer number 0, 1, 2, 3, etc.

The mathematics of standing waves

Two travelling waves moving in opposite directions can be represented by the equations below:

- $y_1 = y_0 \sin(kx - \omega t)$

and

- $y_2 = y_0 \sin(kx + \omega t)$

where

y_0 is the amplitude of the wave,

ω is the angular frequency measured in radians per second (we could use $2\pi f$ instead),

k is equal to $2\pi / \lambda$ (as seen in the travelling wave equations discussed earlier)

x and t are variables for position and time, respectively.

So the resultant wave y equation will be the sum of y_1 and y_2:

- $y = y_0 \sin(kx - \omega t) + y_2 = y_0 \sin(kx + \omega t)$

We can use a trigonometric identity to simplify this to:

- $y = 2y_0 \cos(\omega t) \sin(kx)$

This equation shows not only that the wave oscillates in time, but has also these oscillations vary in the x direction. That is as you move further along the wave (in the x direction) the oscillations vary. Several maxima occur at $x = \lambda/4, 3\lambda/4, 5\lambda/4$, these are the antinodes. Where as at $x = 0, \lambda/2, \lambda, 3\lambda/2$, the function is zero and so the amplitude is always zero – these are the nodes.

Wavelength and the length of string

A standing wave happens if the distance for a wave to travel in a complete circuit, from one point to one end, back to the other end and finally back to where it started, is a whole number of wavelengths. We get stationary waves if the length of the string is a whole number of half wavelengths, i.e. we get stationary waves if

- $n = \dfrac{2L}{\lambda}$

where L is the length of the string and n is an integer, or when

- $\lambda = \dfrac{2L}{n}$

© MoE: FDR Ethiopia Not to be republished

Using velocity $v = f\lambda$ we can show that we get standing waves on the string when

- $\frac{v}{f} = \frac{2L}{n}$

or

- $f = n\frac{v}{2L}$

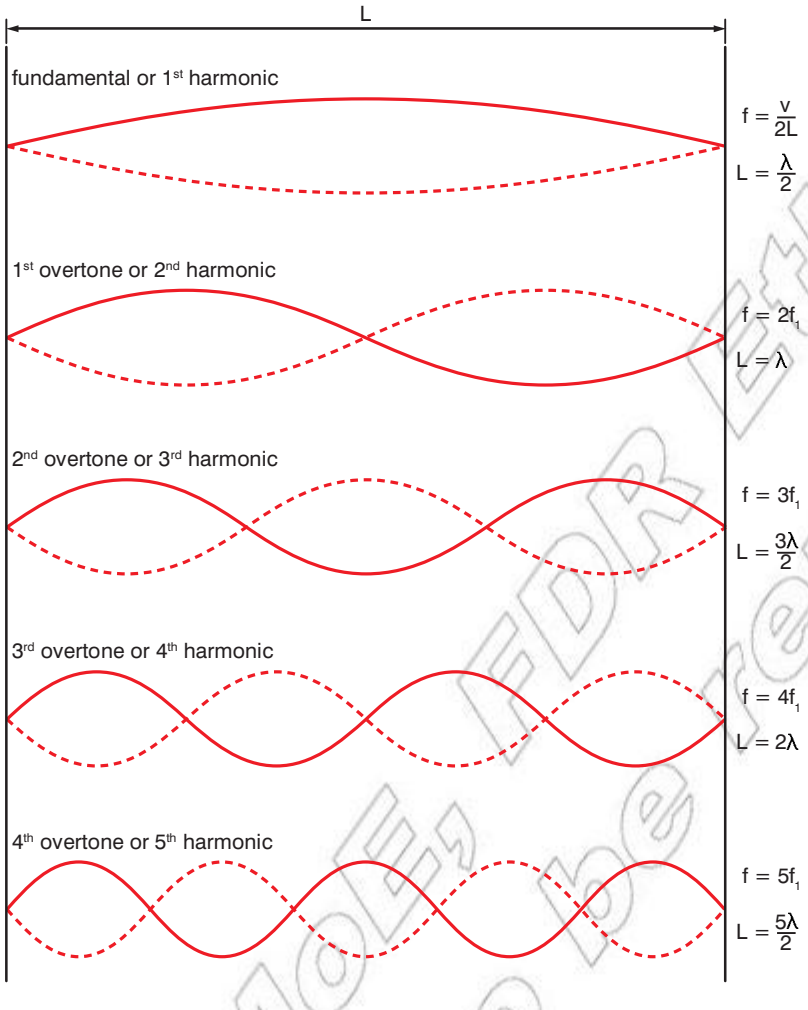


Figure 2.44 *Modes of vibration of a string fixed at both ends*

Figure 2.44 shows the standing waves that can be formed on a string fixed at both ends for values of n from 1 to 5. The lowest frequency of oscillation, when $\lambda = 2L$ and there is just one antinode is known as the fundamental frequency, f_1. All the other possible oscillation frequencies are integer multiples of this fundamental frequency and are known as **harmonics**. For example, the oscillation at $2f_1$ is known as the second harmonic, that at $3f_1$ as the third harmonic, and so on. The first harmonic is the fundamental frequency. The harmonic number is the same as the number of antinodes. A string does not necessarily oscillate at only one of these frequencies; it can oscillate at several different harmonic frequencies at the same time. The resulting shape of the oscillating string can look quite complex.

Worked example 2.5

What frequencies can a string vibrate at if it is 30 cm long and the velocity of travelling waves along the string is 120 m/s?

The fundamental frequency is given by:

- $f_1 = n\frac{v}{2L}$
- $f_1 = \frac{120}{2 \times 0.3}$
- $f_1 = 200$ Hz

Therefore the string can vibrate at 200 Hz, 400 Hz, 600 Hz, 800 Hz, 1000 Hz, 1200 Hz, etc.

KEY WORDS

harmonics *standing waves for which frequencies are integer multiples of the fundamental frequency*

© MoE FDR Ethiopia Not to be republished

Think about this...

The number of antinodes may be used to quickly determine the harmonic. The second harmonic has two antinodes, the third harmonic three, etc.



Figure 2.45 *Guitar strings have different thicknesses (and so mass per unit length is different) plus their tension may be altered to produce different notes.*

DID YOU KNOW?

Strings or parts of strings on a string instrument may resonate at their fundamental or harmonic frequencies when other strings are sounded. For example, an A string at 440 Hz will cause an E string at 330 Hz to resonate, because they share an overtone of 1320 Hz (3rd harmonic of A and 4th harmonic of E).

The fundamental frequency of a string is clearly determined by its length, but it also depends on the velocity at which travelling waves travel along the string. As we have already shown this velocity is given by

- $v = \sqrt{\frac{T}{\mu}}$

where T is the tension in the string and μ is the mass per unit length. Hence, the fundamental frequency of a string is given by

- $f_1 = \frac{1}{2L}\sqrt{\frac{T}{\mu}}$

and so we can make a string produce a higher note if we make it shorter, increase the tension or replace it with a lighter one. These are the parameters that affect the fundamental frequency that the string in a musical instrument produces, but the tone, what makes one instrument sound so different from another arises from the harmonics that are produced at the same time as the fundamental. If a string is tuned to the musical note A (above middle C), then this means that it's fundamental frequency is 440 Hz. But it will also be producing sound waves at the harmonic frequencies 880 Hz, 1320 Hz, 1760 Hz, etc. It is the relative amplitude of these harmonics that determines the tone of the note, and this depends on the detailed design of the musical instrument and how the string is plucked or bowed.

These frequencies of oscillation can be thought of as resonant frequencies. If a sound wave hits the string at one of these frequencies, perhaps coming from another string, it will start to vibrate at that frequency. This is called a sympathetic vibration. When we cause a string to vibrate at a particular resonant frequency, we say that we have excited the corresponding resonant mode of vibration. A key difference from oscillators like pendulums and mass-spring systems is that while those oscillators have just one resonant frequency, a string has multiple resonant frequencies: its fundamental and harmonic frequencies.

Standing waves in organ pipes

Wind instruments generate standing waves in the column of air inside them. A full analysis of the vibration of the air inside a wind instrument is much more complicated than for a string. The vibrating body of air in wind instruments can be varied and complex in shape, but we can arrive at some very good approximations if we stick to straight narrow pipes.

A simple pipe will need to be open at one end, but could be open or closed at the other. Where the end of the tube is closed there is no net oscillation and the sound wave is reflected with a phase shift of π radians. There will always be a node at a closed end. Where the end of the tube is open there is a large net oscillation and the sound wave is reflected with no phase shift. There will always be an antinode at the open end. Pipes with one end closed behave differently from those with both ends open.

Not to be republished © Siyavula Education

If one end of the pipe is closed, and the other open, we will get a node at one end and an antinode at the other. This creates two differences from standing waves on a string. The fundamental occurs when the length of the pipe is just a quarter of a wavelength, and only odd harmonics occur, as shown in Figure 2.46. These diagrams could be slightly misleading as they appear to show the waves as transverse. The vibration of air in the pipe is actually quite complex and certainly not simply transverse: diagrams should be taken as showing amplitude but not direction of oscillation.

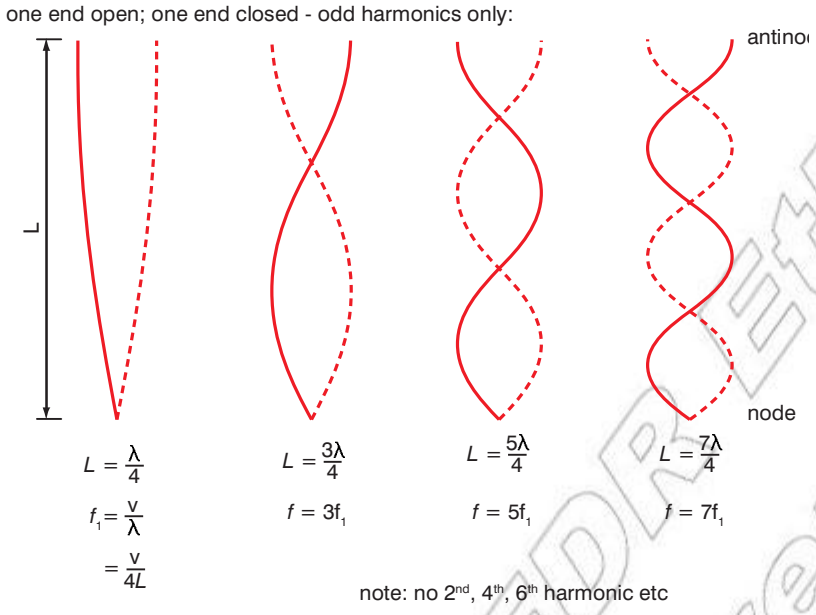


Figure 2.46 *Resonant modes of vibration for air in a narrow column with one end closed and one end open. The diagrams show amplitude of oscillation; they should not be taken as implying that the waves are simple transverse waves.*

For the nth harmonic: $f_n = nv/4L$ where $n = 1, 3, 5$, etc.

If both ends are open, then we get antinodes at both ends. Apart from having antinodes at the ends rather than nodes, this produces the same rules for frequencies and harmonics as a string.

The fundamental occurs for $L = \frac{\lambda}{2}$, and all harmonics, odd and even, can be produced.

Standing waves can also be created using sound in open air. If two loudspeakers are set up, facing each other and some distance apart, playing the same single tone then we have two travelling waves of the same wavelength travelling in opposite directions. This creates a standing wave, but if there is any perfect node it can only be at the centre point between the two speakers, if they are identical and producing the same amplitude.

DID YOU KNOW?

A drum skin has different resonant modes of vibration, just as a string does but in two dimensions rather than one. Where you hit the drum affects which modes are excited and hence the sound the drum produces.

Worked example 2.6

A narrow pipe is 20 cm long and is open at the top and closed at the bottom. Given that the speed of sound is 340 m/s, what frequency sounds might it be possible to produce by blowing across the top?

The wavelength at the fundamental frequency is four times the length of the pipe and therefore

- $\lambda = 0.8$ m

This corresponds to frequency of

- $f = \frac{v}{\lambda} = \frac{340}{0.8} = 425$ Hz

and therefore it might be possible to produce sounds at odd multiples of 425 Hz, i.e. at 425 Hz, 1275 Hz, 2125 Hz, etc.

© NSE EDRETHEA Not to be reproduced

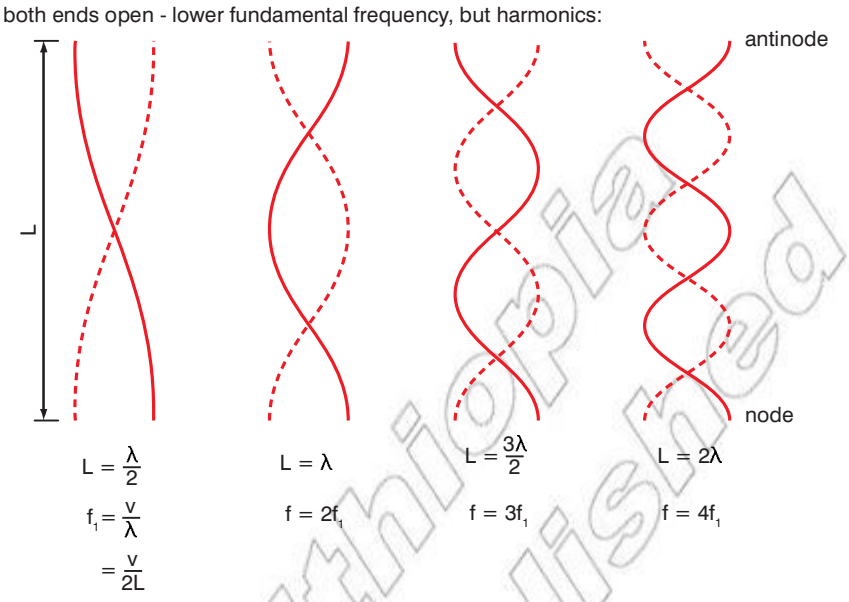


Figure 2.47 *Resonant modes of vibration for air in a narrow column with both ends open. The diagrams show amplitude of oscillation; they should not be taken as implying that the waves are simple transverse wave.*

For the *n*th harmonic: $f_n = nv/2L$ where $n = 1, 2, 3$, etc.

At any other point the sound from one speaker will be louder than from the other and, although a local minimum amplitude will be produced, the complete cancellation to produce a perfect node cannot happen. Standing waves can also be created by directing sound from a single speaker towards a wall, along a path perpendicular to the wall. The wall reflects the sound wave and this produces a standing wave but, as the sound spreads out as it travels, the amplitudes of incident and reflected waves are only similar close to the wall.

DID YOU KNOW?

Reflections of microwaves off building, mountainsides, etc., set up standing wave patterns with nodes or places where the signal strength is very weak. This is one of the biggest problems to be overcome in making mobile phones work.

DID YOU KNOW?

Musicians use the phenomenon of beats to help them tune instruments. If two instruments playing the same note are slightly out of tune, they can hear the beats. If they adjust one of the instruments closer in frequency to the other the beats become slower and slower and when the beat disappears altogether the musicians know the notes really are the same.

Beats

If two notes, differing in frequency by a few hertz are played at similar amplitudes, the phenomenon of beats can be heard. The resulting sound will be at the average of the two frequencies, but it will get louder and quieter at a frequency equal to the difference between the two frequencies. This can be predicted mathematically. If we add two sine waves at frequencies $f + \Delta f$ and $f - \Delta f$, we obtain:

- $\sin(2\pi(f - \Delta f)t) + \sin(2\pi(f + \Delta f)t) = 2\cos(2\pi\Delta f\, t)\sin(2\pi ft)$

which describes a sine wave, $\sin(2\pi ft)$, whose slowly varying amplitude is given by $2\cos(2\pi\Delta f\, t)$, where Δf is half of the frequency difference between the two sine waves. The result is shown in Figure 2.48. The time between nulls, instants of no sound, is half of the period of $\cos(2\pi\Delta f\, t)$, and is hence the time period corresponding to the frequency difference. In other words, if we play two notes at frequencies f_1 and f_2 the resulting sound has a beat frequency of

- $f_B = |f_2 - f_1| = \frac{1}{T_B}$

For Ethiopia to be republished

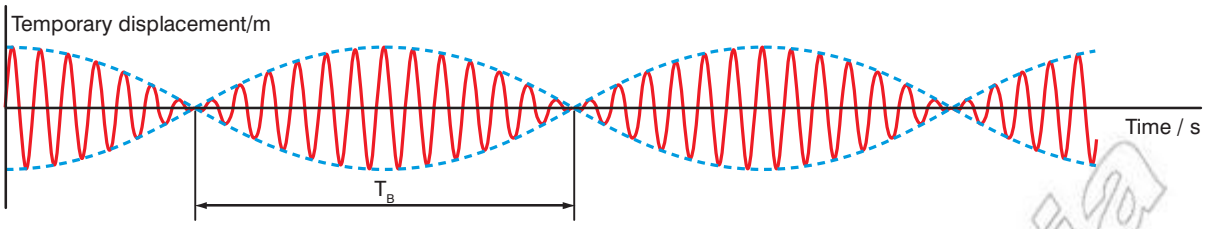


Figure 2.48 *Beats produced by the addition of two sine waves of the same amplitude but slightly different frequencies. The resultant sound is at the average of the two frequencies (red sine wave) but with an amplitude that increases and decreases at a "beat frequency" that is the difference between the two frequencies.*

Summary

In this section you have learnt that:

- A travelling wave transfers energy (and information) with no permanent movement of mass.
- A transverse wave is one in which the oscillation is perpendicular to the direction of propagation.
- A longitudinal wave is one in which the oscillation is along (parallel to) the direction of propagation.
- Wave speed $v = f\lambda$.
- The travelling wave function is $Y = A\sin\left(2\pi\frac{x}{\lambda} - 2\pi ft\right)$.
- The principle of superposition states that if two or more waves pass through a single point then the resultant instantaneous displacement at that point is the sum of the displacements that would be created separately by each wave.
- The phase difference between two oscillations is the angular difference in their timing, e.g. a half cycle difference is a phase difference of π radians.
- Constructive interference is the production of large oscillations by superposition of oscillations due to waves in phase with each other.
- Destructive interference occurs when oscillations due to waves are in antiphase with each other, and completely or partially cancel each other out.
- Nodes are points in a standing wave where no oscillation occurs.
- Antinodes are points in standing waves where the amplitude of oscillation is at a maximum.
- The distance between nodes in a standing wave is $\frac{\lambda}{2}$.

Ethiopia published

- At a fixed end of a string or closed end of an air column, travelling waves are reflected with a phase shift of π radians, giving rise to a node in the standing wave produced.
- At a free end of a string or open end of an air column, travelling waves are reflected with zero phase shift, giving rise to an antinode in the standing wave produced.
- The fundamental mode of vibration of a string or air column is the standing wave at the lowest possible frequency.
- Harmonics are integer multiples of the fundamental frequency
- A string with fixed ends supports standing waves such that $n\frac{\lambda}{2} = L$, where L is the length of the string and n is an integer.
- For a pipe with one closed end and one open end the fundamental mode of vibration occurs for $n\frac{\lambda}{4} = L$, and only odd harmonics can occur.

Review questions

1. If the speed of sound in air is 340 m/s, what is the wavelength of a sound wave at 512 Hz?
2. What is the frequency of red light with a wavelength (in free space) of 630 nm?
3. Two sinusoidal waves both have a frequency of 200 Hz. The amplitude of one is 1 cm, the amplitude of the other is 2 cm. Sketch graphs of displacement against time for the oscillations produced by each wave separately and their resultant at a point P where they cross, if
   a) they arrive at P in phase, and
   b) they arrive at P in antiphase.
4. A travelling wave on a string, of amplitude 2 mm, frequency 500 Hz and speed 300 m/s, can be described by the function
   $Y = A\sin\left(2\pi\frac{x}{\lambda} - 2\pi ft\right)$
   a) Sketch graphs of displacement Y against distance x for this wave, for the first 1.2 m:
      i) for time $t = 0$, and
      ii) for time $t = 0.5$ ms
   b) Sketch graphs of displacement Y against time t for the oscillation produced by this wave for the first 4 ms
      i) at the source where $x = 0$, and
      ii) at a distance $x = 30$ cm from the source.

© MoE, EDB Not to be reproduced

5. A violin string is 32 cm long and has a mass per unit length of 2 g/m. What tension is required for the string to produce an A, i.e. for its fundamental frequency to be 440 Hz?
6. Imagine that you swinging backwards and forwards on a child's swing and you are listening to music coming from a loudspeaker in front of you. Explain why the music might not sound right.
7. A pipe, 68 cm long, is open at one end and closed at the other. When air is blown across the open end sound is produced at 110 Hz.
   a) What is the velocity of sound along the pipe?
   b) Blowing harder will produce a higher note. What is the next frequency that the pipe can produce?
8. In an experiment to measure the speed of sound in air, a speaker directs sound towards a wall, along a path perpendicular to the wall. The wall reflects the sound wave and this produces a standing wave. A microphone and electronic measuring device is used to measure the amplitude of the sound at different distances from the wall. Minimum values of amplitude are detected at 28 cm when the frequency used is 600 Hz.
   a) What is the measured speed of sound?
   b) Explain why the minimum values of sound are not zero and why they are more difficult to detect further from the wall.

2.3 Sound, loudness and the human ear

By the end of this section you should be able to:

- Define the intensity of sound and state the relationship between intensity and distance from the source.
- Describe the dependence of the speed of sound on the bulk modulus and density of the medium. Use $v = \sqrt{B/\rho}$
- Give intensity of sound in decibels, and define the terms threshold of pain and threshold of hearing.
- Describe the intensity level versus frequency graph to know which the human ear is most sensitive to.
- Analyse resonance conditions in air columns in quantitative terms.
- Explain the Doppler effect, and predict in qualitative terms the frequency change that will occur in a variety of conditions.
- Describe some practical applications of the Doppler Effect.

EDR Ethiopia republished

KEY WORDS

loudness *the audible strength of a sound, which depends on the amplitude of the sound wave*

intensity *the energy received by each square metre of a surface per second*

power per unit area *the power received by a square metre of a surface*

Sound loudness and intensity

The **loudness** of sound is difficult to measure scientifically. How loud a sound appears depends very on the listener, it is quite subjective. In general, the louder the sound, the greater its **intensity**.

Intensity can be defined precisely. Intensity at a point is then energy flowing through a unit area (1 m^2) per unit time (1 s). We know that the energy in an oscillation is equal to the kinetic energy at the equilibrium position

$= \frac{1}{2}mv_0^2 = \frac{1}{2}m\omega^2 A^2$.

The mass of air with a cross section of 1 m^2 disturbed by a plane sound wave each second is the density of air ρ times the volume disturbed, and the volume disturbed is 1 m^2 time the distance travelled by the wave each second (its velocity). Hence, the intensity of a sound wave is given by

$I = \frac{1}{2}\rho v\omega^2 A^2$,

where ρ is the density of the air, v is the velocity of sound, $\omega = 2\pi f$ where f is the frequency and A is the amplitude of oscillation. Hence we see that intensity, measured in W/m^2, is proportional to amplitude squared and to frequency squared.

Sound intensity is defined as the sound **power per unit area**. The usual context is the measurement of sound intensity in the air at a listener's location. The basic units are watts per m^2 or W/m^2.

Alternatively the intensity of sound from source can be calculated assuming the sound spreads out equally in all directions. You may recall the intensity from a point source is given by:

- $I = \frac{P}{A}$

where

P = power of the source in W

A = area through which the sound is transmitted.

If we assume the sound travels out equally in all directions then the area covered is equal to the surface area of a sphere. So the equation becomes:

- $I = \frac{P}{A} = \frac{P}{4\pi r^2}$

where r is the distance from the source in m.

You may recall from Grade 9 this is an inverse square relationship, is you double the distance the sound intensity falls by 4 (2^2).

© MoE: EBS Ethiopia
Not to be republished

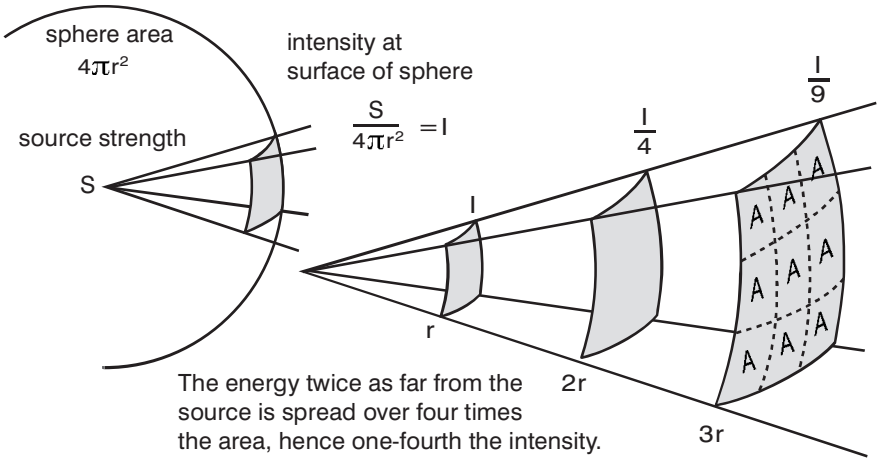


Figure 2.49 *The intensity from the source varies as an inverse square relationship.*

Worked example 2.7

A large explosion is detected 200 m away. The intensity at this distance is measured to be 400 MW/m^2. A few seconds later the explosion is recorded 3.2 km away. Find the power of the original explosion and the intensity recorded at the larger distance.

- $I = \frac{P}{A} = \frac{P}{4\pi r^2}$ — *State the relationship to be used*
- $P = IA = I \times 4\pi r^2$ — *Rearrange to make P the subject*
- $P = 400 \times 10^6 \times 4 \times \pi \times 200^2$ — *Substitute known values*
- $P = 2.0 \times 10^{14}$ W — *Solve for P and give the units*

We can now use the intensity equation to determine the intensity at 3.2 km.

- $I = \frac{P}{A} = \frac{P}{4\pi r^2}$ — *State the relationship to be used*
- $I = 2.0 \times 10^{14} / (4 \times \pi \times 3200^2)$ — *Substitute known values*
- $I = 1.6 \times 10^6$ W/m^2 — *Solve for* I *and give the units*

© Ethiopia Not published

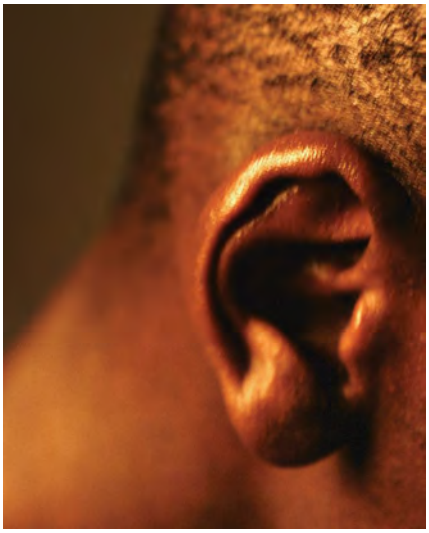


Figure 2.50 *The human ear can detect tiny changes in pressure.*

Hearing and the decibel

Many sound intensity measurements are made relative to the **threshold of hearing** intensity I_o.

This is defined as:

- $I_o = 1 \times 10^{-12}$ W/m^2

One common way to measure the loudness of sound is to use the **decibel** scale. The decibel (dB) is a logarithmic unit of measurement. Decibels measure the ratio of a given intensity I to the threshold of hearing intensity I_o. This means so that I_o has the value 0 decibels (0 dB).

The intensity of a sound in dB is given by:

- $I(dB) = 10\log_{10}\left(\frac{I}{I_o}\right)$

The human ear is incredibly sensitive to sound. I_o represents a pressure change of less than one billionth of standard atmospheric pressure. In reality the actual threshold of the average human is closer to 2.5×10^{-12} W/m^2. This corresponds to 4 dB.

The threshold of hearing varies with frequency. The ear is most sensitive to sounds between 2000 and 5000 Hz. This can be seen on a **hearing curve**.

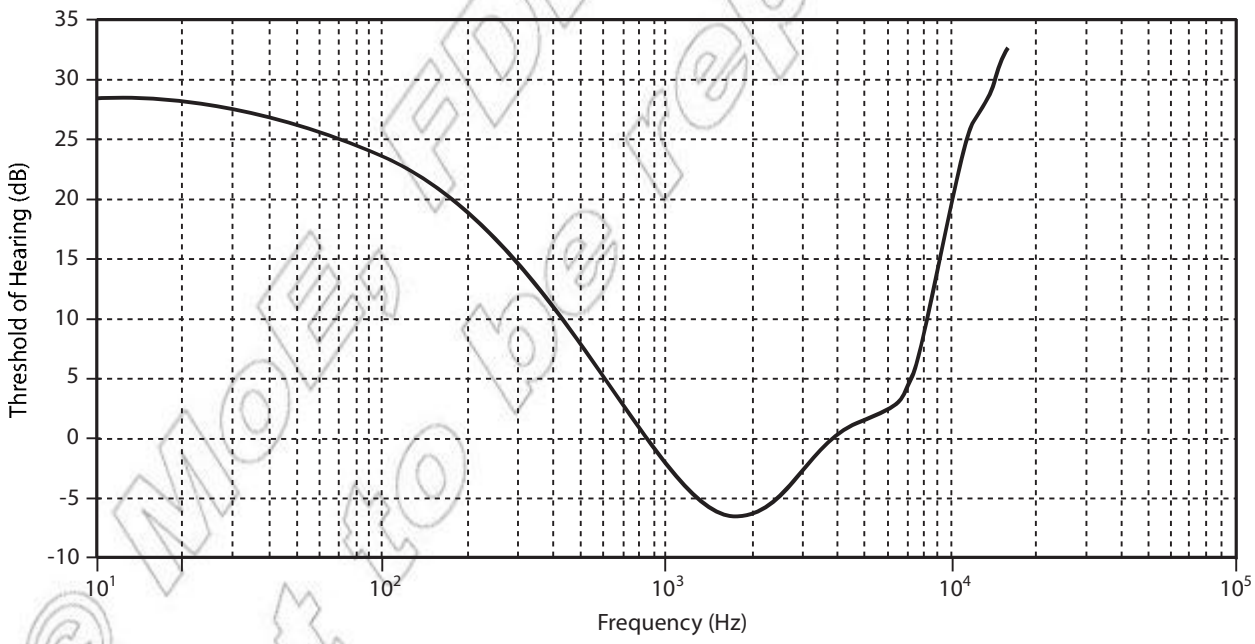


Figure 2.51 *A hearing curve shows how the threshold of hearing varies with frequency.*

This curve illustrates why certain sounds at the same intensity appear to have different volumes. The human ear is simply better a detecting some frequencies of sound than others. The exact shape of the curve depends on a number of factors including age, exposure to loud sounds and the physical characteristics of the ear.

The upper limit of human hearing is also rather subjective. A common upper limit is the **threshold of pain**. This is the point at

KEY WORDS

threshold of hearing *a standard threshold against which measurements of sound intensity are made. It is equal to 0 dB intensity.*

decibel *logarithmic unit of measurement of the loudness of sound*

© Mofed Education Not to be republished

which pain begins to be felt by the listener. This value varies from individual to individual and for a given individual over time. In general, younger people are more tolerant of loud sounds. A common value for this limit is 130 dB.

One way to express the range of human hearing is to use the standard threshold of hearing up to the threshold of pain. This represents a huge range, from I_o to $1 \times 10^{13}\ I_o$!

The speed of sound

As discussed in the previous chapter the speed of sound depends on the medium the sound is travelling through. If the sound is travelling through a solid the speed is given by:

- $v = \sqrt{\frac{Y}{\rho}}$

where

Y = Young's modulus of the solid (effectively a measure of the stiffness of the solid) in Pa

ρ = density of the solid in kg/m^3

If the sound is travelling through a fluid (liquid or gas) there is a similar equation:

- $v = \sqrt{\frac{B}{\rho}}$

where

B = bulk modulus of the fluid (effectively a measure of the compressibility of the fluid) in Pa

ρ = density of the fluid in kg/m^3

Both equations show the speed of sound increases with the 'stiffness' of the material and decreases with the density. The table below gives typical values for the Young's/bulk modulus and densities of materials. The actual values vary depending of the exact composition of the substance.

Material	Y (GPa)	B (Pa)	ρ (kg/m^3)
Air	–	1.0×10^5	1.0
Water	–	2.2×10^9	1000
Glass	40	–	4000
Steel	160	–	7500
Diamond	442	–	3500

KEY WORDS

hearing curve *a graph which shows how threshold of hearing varies with frequency*

threshold of pain *the point at which pain, caused by sound, begins to be felt by the listener*

DID YOU KNOW?

You may recall that the speed of sound through air is given by the equation

$$v = 331\sqrt{1 + \frac{T_k}{273°\,C}}\ \text{m/s}$$

Think about this...

Just like a transverse travelling wave a longitudinal travelling wave (like sound) can be represented using a sine or cosine function. In this case.

$$S(x, t) = A\cos\left(\frac{2\pi x}{\lambda} - 2\pi f t\right)$$

In this equation S is used instead of y to denote horizontal displacement of a particle away from its equilibrium position.

NOT FOR DISTRIBUTION



Figure 2.52 *You hear the Doppler effect whenever a siren travels passed.*

DID YOU KNOW?

The Doppler effect (or Doppler shift), is named after Austrian physicist Christian Doppler who first explained it in 1842. The Doppler effect works with light too. Analysing the light coming from distant galaxies has shown us that they are moving away from us and allowed us to calculate how fast.

Discussion activity

If you stood in the middle of a road with a blindfold on and could hear the siren of an ambulance coming towards you, you should be much more worried if the pitch did not change at all than if the pitch started to drop a little as the sound get louder. Why?

The Doppler effect

When a vehicle with siren, such as an ambulance or police car, goes past us we notice a pronounced change in pitch. As the siren approaches the pitch is higher; as it moves away from us the pitch is lower. This is known as the **Doppler effect**, or Doppler shift.

If a sound source transmits sound in all directions, like a siren, and it is stationary, then the wavelength of the transmitted sound, and hence its frequency, is same in all directions, and this is the frequency of the sound source. However, if the source is approaching a listener the sound waves are compressed. This is because, after emitting one wave front, the source moves towards the listener and emits the next wave front from a position closer to the listener. The wavelength of the sound arriving at the listener is made shorter and, since the speed at which the sound travels, the frequency, calculated by $f = v/\lambda$ is higher. The faster the source is moving the bigger the change, or shift, in frequency. If the source is moving away from the listener, the opposite happens: the wavelength is made longer and the frequency made lower. This is shown in Figure 2.53. Exactly the same effect is observed if the source remains stationary and the listener moves towards or away from the source; it is the relative velocity that matters. The increase in wavelength, $\Delta\lambda$, is given by

- $$\frac{\Delta\lambda}{\lambda} = \frac{\text{relative velocity of listener and source away from each other}}{\text{velocity of sound}}$$

Worked example 2.8

You measure the frequency you hear from an ambulance siren at 466 Hz. You know that the siren actually transmits sound at a frequency of 440 Hz, and the speed of sound in air is 340 m/s. What is the velocity of the ambulance relative to you?

For f = 440 Hz the wavelength is

- $$\lambda = \frac{v}{f} = \frac{340}{440} = 0.7727 \text{ m}$$

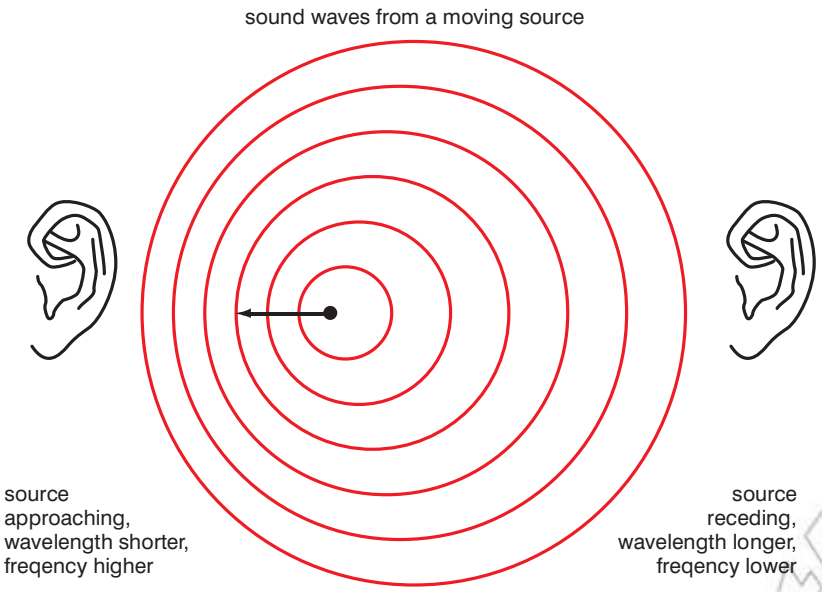
For f = 466 Hz the wavelength is

- $$\lambda = \frac{v}{f} = \frac{340}{466} = 0.7296 \text{ m}$$

Therefore $\Delta\lambda = 0.7296 - 0.7727 = -0.0431$ m, and the velocity of the ambulance is

- $$\frac{\Delta\lambda}{\lambda} \times \text{speed of sound} = \frac{-0.0431}{0.7727} \times 340 = -19.0 \text{ m/s}$$

i.e. the ambulance is approaching at 19.0 m/s (= 68 km/h[1]).



KEY WORDS

Doppler effect *a change in the observed frequency of a wave occurring when the source and observer are in motion relative to each other, with the frequency increasing when the source and observer approach each other and decreasing when they move apart*

Figure 2.53 *The Doppler effect. Sound waves from an approaching source are compressed and therefore shorter, giving a higher frequency sound; waves from a receding source are stretched and therefore longer, giving a lower frequency sound.*

There are three situations to consider:

Sound source is stationary relative to the listener

In this case the frequency received by the listener is the same as that produced by the source.

- $f_L = f_s$

Sound source is moving toward the listener (or vice-versa)

In this case the frequency received by the listener greater than the frequency produced by the source. The relationship is given by

- $f_L = f_s \frac{1}{1-\frac{v_s}{v}}$

where

v_s = speed of source

v = speed of sound in air

Sound source is moving away from the listener (or vice-versa)

In this case the frequency received by the listener lower than the frequency produced by the source. The relationship is given by

- $f_L = f_s \frac{1}{1+\frac{v_s}{v}}$

© MoE FDR Ethiopia 2022 Not to be republished

Applications of Doppler effect

The Doppler effect has a number of applications including:

Astronomy

Observations of the spectral lines in the visible spectrum of light from distant galaxies show a red-shift. This has been used to demonstrate the universe is expanding and is a key piece of evidence in support of the big bang theory. The Doppler effect is used to measure the speed at which stars and galaxies are approaching or receding from us.

Medical imaging and blood flow measurement

An echocardiogram is used to determine the direction and velocity of blood flow using the Doppler effect (in this case ultrasound is used).



Figure 2.54 *Most types of radar use the Doppler effect*

Other flow measurements

Instruments like the laser Doppler velocimeter are used to measure velocities in a fluid flow. In this case a laser light is fired at a moving fluid. A Doppler shift is observed from reflections off of particles moving with the fluid.

Radar

The Doppler effect is used in some types of radar. It is used to measure the velocity of a range of objects. A radar beam is fired at a moving target and reflects from the surface back to the detector. Any change in wavelength is then recorded and the object's velocity can be accurately determined. Doppler radar is used in a range of applications, including the speed of motorist, tennis serves, even the speed of a football struck towards a goal.

Summary

In this section you have learnt that:

- Intensity of sound, measured in W/m^2, is proportional to amplitude squared and to frequency squared.
- Sound intensity may be found by:

  $I = \frac{1}{2}\rho v\omega^2 A^2$ or $I = \frac{P}{A} = \frac{P}{4\pi r^2}$
- The threshold of hearing intensity $I_o = 1 \times 10^{-12}$ W/m^2.
- The intensity of a sound in dB is given by:

  $I(dB) = 10\log_{10}\left(\frac{I}{I_o}\right)$.
- One way to express the range of human hearing is to use the standard threshold of hearing up to the threshold of pain.
- The speed of sound through a solid is given by $v = \sqrt{\frac{Y}{\rho}}$.

© MoE-EDP Ethiopia Not to be republished

- The speed of sound through a fluid is given by $v = \sqrt{\frac{B}{\rho}}$.
- The Doppler effect causes an increase in the frequency of the received sound wave if the sound source and listener are moving towards each other, a decrease in frequency if they are moving away from each other.
- When the source of sound is moving towards the listener $f_L = f_s \frac{1}{1-\frac{v_s}{v}}$
- When the source of sound is moving away from the listener $f_L = f_s \frac{1}{1+\frac{v_s}{v}}$

Review questions

1. Calculate the intensity in dB of a sound with an intensity of 6.2×10^{-6} W/m^2.
2. Determine the intensity of the threshold of pain for an average person.
3. Calculate the speed of sound through:
   a) water
   b) steel
   c) diamond.
4. Imagine that you swinging backwards and forwards on a child's swing and you are listening to music coming from a loudspeaker in front of you. Explain why the music might not sound right.
5. The horn of a stationary car emits sound at a frequency of 440 Hz. What frequency of note will you hear if you drive towards this car at 20 m/s? (The speed of sound in air = 340 m/s.)
6. Two cars drive along the same road towards each other, one at 15 m/s and the other at 12 m/s. Each car horn sounds at 256 Hz. Calculate the frequency that the driver of each car hears coming from the other car.
7. Describe three uses of the Doppler effect.

End of unit questions

1. A simple pendulum is made from a bob of mass 0.040 kg suspended on a light string of length 1.4 m. Keeping the string taut, the pendulum is pulled to one side until it has gained a height of 0.10 m. Calculate
   a) the total energy of the oscillation
   b) the amplitude of the resulting oscillations
   c) the period of the resulting oscillations

© FDR Ethiopia Not to be republished

d) the maximum velocity of the bob

e) the maximum kinetic energy of the bob.

2. A piston in a car engine has a mass of 0.75 kg and moves with motion which is approximately simple harmonic. If the amplitude of this oscillation is 10 cm and the maximum safe operating speed of the engine is 6000 revolutions per minute, calculate:

   a) maximum acceleration of the piston

   b) maximum speed of the piston

   c) the maximum force acting on the piston.

3. An experiment is carried out to measure the spring constant of a spring. A mass of 500 g is suspended on the spring. It is pulled down a small distance and the time for 20 oscillations is measured to be 34 s.

   a) Explain why the mass performs simple harmonic motion.

   b) What is the spring constant?

   c) What is the equilibrium extension of the spring?

   d) If the mass and spring were to be moved to the surface of the Moon (where the gravitational field strength in 1.6 N/kg) what would the effect be on the time period of oscillation and on the equilibrium extension of the spring?

4. A car of mass 820 kg has an under damped suspension system. When it is driven by a driver of mass 80 kg over a long series of speed bumps 10 m apart at a speed of 3 m/s the car bounces up and down with surprisingly large amplitude.

   a) Explain why this effect occurs.

   b) Calculate the net spring constant of the car suspension system.

5. If you are given a metal rod and a hammer, how must you hit the rod to produce:

   a) a transverse wave, and

   b) a longitudinal wave?

6. A string of a musical instrument has a fundamental frequency of 196 Hz. What are the frequencies of the 2^{nd}, 3^{rd} and 4^{th} harmonics of this string.

7. A string is 1.6 m long, and waves travel along it at 2400 m/s.

   a) Sketch a labelled diagram to describe the stationary wave pattern for 4^{th} harmonic mode of vibration.

   b) Calculate the frequency of this vibration.

   c) On the same set of axes, sketch graphs of displacement against time for the oscillations 0.2 m and 0.5 m from one end of the string.

© MoE, FDR Ethiopia
Not to be republished

8. When tuning a piano, a musician plays a note that should be at 110 Hz while at the same time tapping a 110 Hz tuning fork and holding it next to the strings. He hears beats at 4 Hz.
   a) State and explain what frequencies the piano could be producing.
   b) Draw a sketch graph to show the resultant sound as a variation in displacement against time. Label the time axis with values.
   c) Explain how the musician now finishes tuning his piano
9. A short string will usually oscillate with smaller amplitude than a longer string. Explain the consequence for the relative loudness of different frequencies played by a string instrument if this was not the case.
10. A whistle producing a sound at 1 KHz is whirled in a horizontal circle at a speed of 18 m/s. What are the highest and lowest frequencies heard by a listener standing a few metres away, if the speed of sound in air is 340 m/s?

© MoE, FDR Ethiopia
Not to be republished

Wave optics

Unit 3

Contents

Section	Learning competencies
3.1 Wave fronts and Huygens's principle (page 109)	• Define the term wave front. • State Huygens's principle.
3.2 Reflection and refraction of plane wave fronts (page 113)	• Understand reflection and refraction of plane wave fronts (including diagrams).
3.3 Proof of laws of reflection and refraction using Huygens's principle (page 116)	• Understand the proof of the laws of reflection and refraction using Huygens's principle. • State the laws of reflection and refraction. • Describe reflection and refraction in terms of the wave nature of light.
3.4 Interference (page 120)	• Describe the phenomena of wave interference as it applies to light in qualitative and quantitative terms using diagrams and sketches. • Compare the destructive and constructive interference of light with superposition along a string. • Identify the interference pattern produced by the diffraction of light through narrow slits (single and double slits). • Define an interferometer as a device which uses the interference of two beams of light to make precise measurements of their path difference. • Define thin film interference and apply and use the equations to solve problems.
3.5 Young's double slit experiment and expression for fringe width (page 128)	• Explain the interference in Young's double slit experiment. • Carry out calculations involving Young's double slit experiment.
3.6 Coherent sources and sustained interference of light (page 131)	• State the conditions necessary for the interference of light to be shown.
3.7 Diffraction due to a single slit and a diffraction grating (page 133)	• Describe the diffraction due to a single slit, including the interference caused rays of light coming from different parts of the slit. • Describe and explain the diffraction of light in quantitative terms using diagrams. • Describe the effects of using a diffraction grating.

In the previous unit we looked at wave motion in general. This unit concentrates on one particular type of wave – light.

The theories on the nature of light are wide ranging. They include the ancient Greek model of light particles swarming from sources, to Leonardo da Vinci's ideas comparing light and sound right up to the modern day ideas of wave–particle duality. Newton was a great physicist. However, his powerful reputation actually impeded the development of the understanding of the true nature of light. Newton proposed that light was made up of tiny particles called corpuscles. Despite the best efforts of some of his contemporaries, his ideas remained in force for hundreds of years, even when there was a significant body of evidence against his theories.

The main alternative theory was proposed by the Dutch scientist Christiaan Huygens. He developed what he called the wave nature of light in his *Treatise on Light*. This theory provides an explanation for reflection and refraction and may be used to verify both the law of reflection and Snell's law.

Light waves exhibit reflection, refraction, diffraction and interference. These are properties of all waves, but due to the tiny nature of the **wavelength** of light these effects have slightly different characteristics and mathematical models used to describe them. This unit will further analyse the wave nature of light. Using Huygens's principle common wave phenomena such as refraction and diffraction will be explained.

3.1 Wave fronts and Huygens's principle

By the end of this section you should be able to:

- Define the term wave front.
- State Huygens's principle.

What are wave fronts?

A **wave front** is an imaginary line joining points of a travelling wave that are in phase. You can think of this as a line joining all points in space that are reached at the same instant by a wave.

Imagine an oscillating source placed in water. As it moves up and down the wave travels out in all directions from the source. This may be seen as the ripples moving out from the centre.

At distance A from the source, all the points have the same displacement due to the water wave. All these points are in phase. Joining these points creates a circle with the source at its centre. This line is a wave front.

The circle joining the displacements at B is an example of another wave front.

KEY WORDS

wave front *an imaginary line joining the points of a travelling wave that are in phase*

wavelength *the minimum distance between identical points on adjacent waves, equal to the distance between wave fronts*

ray diagrams *where the direction of a wave is represented by a line*

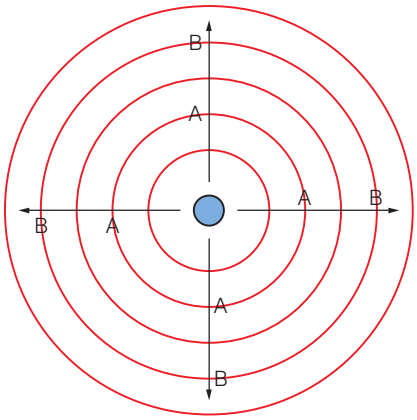


Figure 3.1 *Wave fronts travelling out from a single source*

© NIED Ethiopia Not to be republished

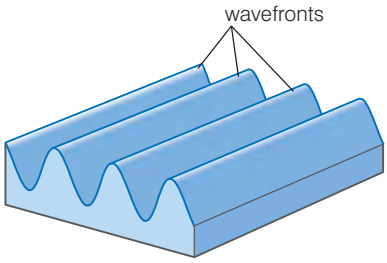


Figure 3.2 *Wave fronts on a plane wave can be thought of as a line along one of the peaks or compressions.*



Figure 3.3 *The peaks of these ripples can be considered to be wave fronts.*

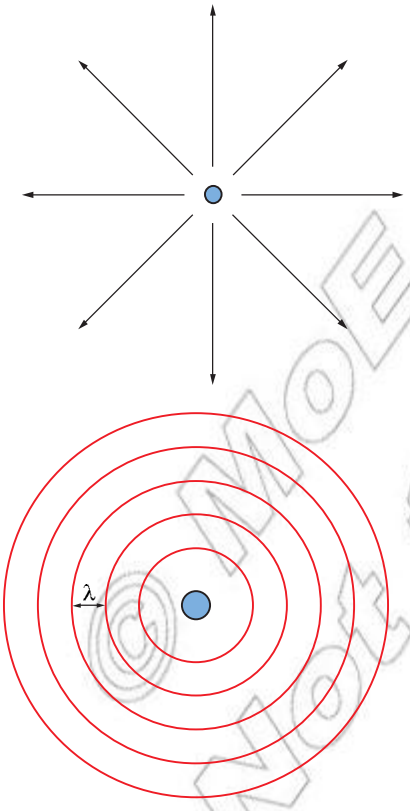


Figure 3.5 *A ray diagram (top) and a wave front diagram (bottom) showing waves travelling out in two dimensions from a source. This might be light from a light bulb.*

Along one particular peak of a transverse wave or one compression of a longitudinal wave the particles are all in phase. This means a wave front can also be thought of as a line joining up one particular peak or compression. As a consequence, the distance between wave fronts is equal to the wavelength of the wave.

Wave fronts and ray diagrams

We have often represented waves using **ray diagrams** where a single line represents the direction of the wave. We have also looked at waves 'side on' to demonstrate amplitude, time period and wavelength. However, there are situations where a wave front diagram proves more useful. These diagrams provide us with a sort of 'top down' view.

Below are two different diagrams to represent plane waves travelling from a source (for example, a stick held horizontally in water and then oscillated vertically).

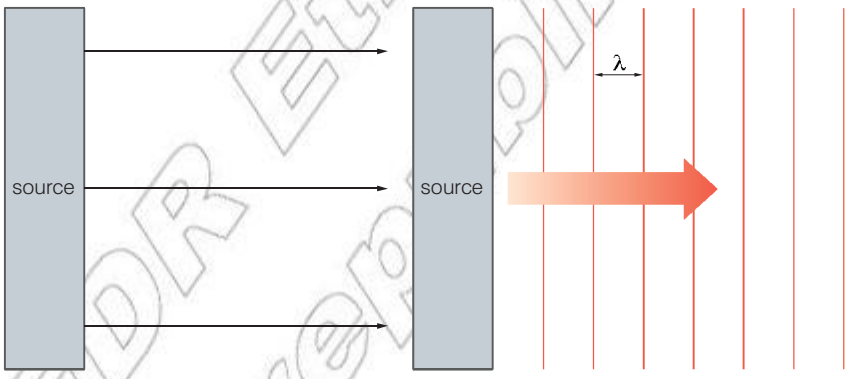


Figure 3.4 *A ray diagram (left) and a wave front diagram (right) showing plane waves travelling out from a source*

Notice that the wave front diagram allows us to represent the wavelength of the wave as the distance between two wave fronts. However, without the orange arrow the direction of the travelling wave would not be known.

The same comparison may be made for waves travelling out in two dimensions.

Both techniques are often combined in a single diagram.

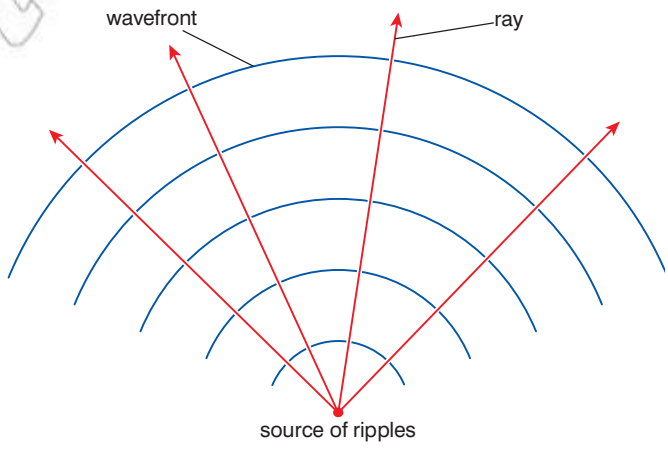


Figure 3.6 *A number of rays at 90° to the wave fronts travelling out from a source.*

© MOE EDR Ethiopia Not to be reproduced

Observing wave fronts with a ripple tank

Ripple tanks are often used to produce images of the wave fronts along a water wave.

Figure 3.8 shows two examples of images obtained using ripple tanks.

Activity 3.1: Ripple tanks

Use a ripple tank to observe wave fronts. Alter the frequency of the oscillations of the source and observe the effect on the wavelength.

It is tempting to assume that the dark area is due to the peak; however, this is not true. The peak acts like a lens and focuses the light underneath it. This means that the bright lines are the peaks and the dark areas are the troughs of the water wave.

Using ripple tanks it is easy to demonstrate and observe the effects of **reflection**, **refraction**, **diffraction** and **interference** on the wave fronts produced by a source.

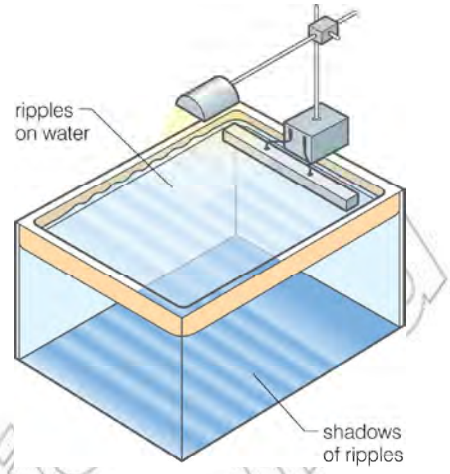


Figure 3.7 *A ripple tank can produce clear images of the wave fronts along water waves.*

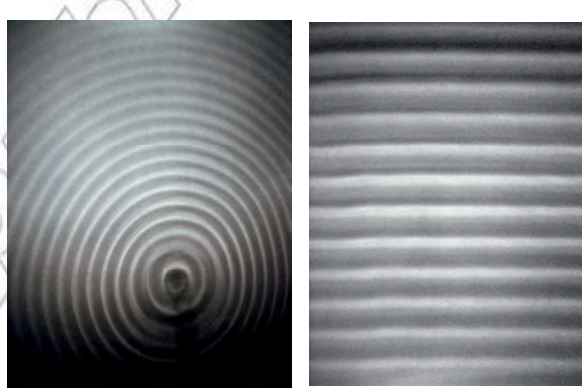


Figure 3.8 *Circular waves and plane waves produced in a ripple tank*

How are wave fronts formed?

The Dutch physicist Christiaan Huygens developed a theory of how wave fronts were formed. It is often referred to as **Huygens's principle** or Huygens' wave construction. Huygens applied his theory to light. He proposed a wave theory of light – in fact he developed his principle in support of his wave theory.

This theory was controversial and was not widely accepted until well over 100 years after his death (more on this later).

Huygens's principle states:

- **Every point on a wave front acts as a source of spherical secondary wavelets.**
- **These secondary wavelets spread out in all directions and have the same frequency and speed as the original wave (and so the same wavelength).**
- **A new wave front is formed as these wavelets combine together.**

Huygens's principle was slightly modified by Jean Fresnel to explain why no back wave was formed and we will use this modification to demonstrate how this principle leads to the formation of a new wave front.



Figure 3.9 *Christiaan Huygens lived from 1629 until 1695.*

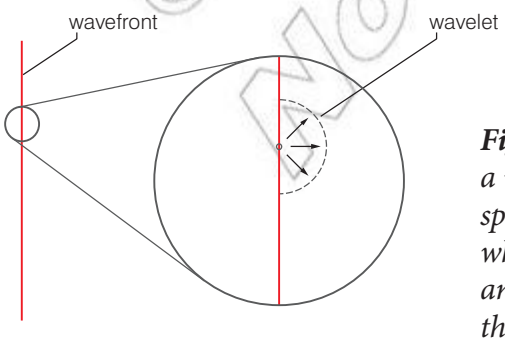


Figure 3.10 *Every point along a wave front acts as a source of spherical secondary* ***wavelets****, which travel at the same speed and in the same direction as the wave.*

KEY WORDS

Huygens's principle *principle describing how waves propagate through a medium*

© NOT FREE TO BE REPRODUCED

DID YOU KNOW?

The Dutch scientist Christiaan Huygens (1629–1695) is not only famous for his wave theory of light. He also discovered more information about the rings of Saturn and he even invented the first useful pendulum clock!

Activity 3.2: Five consecutive wave fronts

Draw a diagram similar to Figure 3.11 to show the formation of five consecutive wave fronts for a travelling wave.

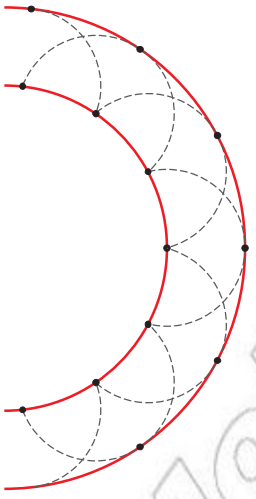


Figure 3.12 *Wavelets from a spherical wave front create another spherical wave front in front of the first.*

If the original wave front is from a plane wave then the wavelets combine to form another plane wave having travelled a distance equal to the product of the wave speed and the time taken ($v\Delta t$).

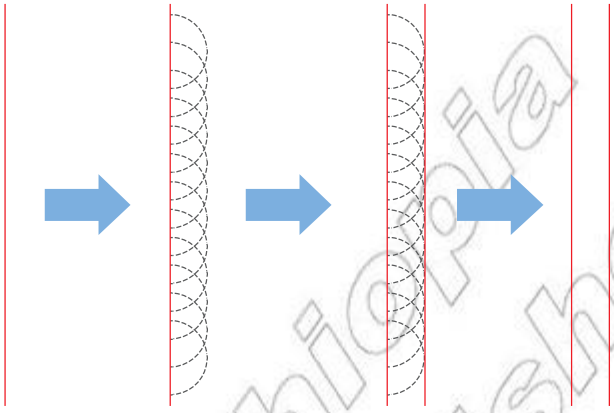


Figure 3.11 *These wavelets combine to form a new wave front parallel to the original one. This process then repeats as the wave moves through space.*

This process may be applied to non-plane wave fronts. If the original wave front was spherical then the new wave front will also be spherical.

In simple terms, Huygens's principle means you can view the 'edge' of the wave as actually creating a series of circular waves. These waves combine together to form a new wave front. In most cases this process just continues the wave propagation as the wave travels through the medium. However, Huygens's principle can also be used to explain important wave effects such as diffraction and refraction (more on this in sections 3.3 and 3.4).

Summary

In this section you have learnt that:

- A wave front is a line joining all parts of a wave that are in phase.
- Huygens's principle provides a description of how waves propagate through a medium. It states that all points along a wave front produce a series of secondary wavelets. These wavelets travel at the same speed and have the same frequency as the original wave. The wavelets combine to form a new wave front and this process continues.

Review questions

1. Define what is meant by the term wave front.
2. State Huygens's principle.
3. Use a series of diagrams and Huygens's principle to demonstrate how circular ripples travel out from a point source.

© MoE-FDR Ethiopia Not to be republished

3.2 Reflection and refraction of plane wave fronts

By the end of this section you should be able to:

- Understand reflection and refraction of plane wave fronts (including diagrams).

Reflection in terms of wave fronts

KEY WORDS

law of reflection *the angle of incidence of a wave equals the angle of reflection*

The diagram in figure demonstrates the **law of reflection**.

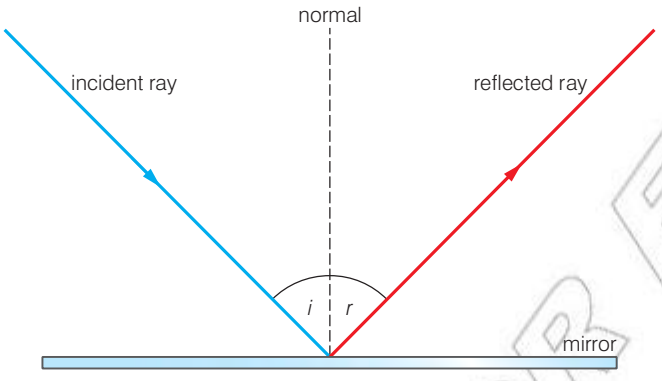


Figure 3.13 *A ray diagram to demonstrate the law of reflection*

All waves obey the law of reflection. This states:

- angle of incidence = angle of reflection
- $i = r$ (or $\theta_1 = \theta_2$)

When constructing a wave front diagram of a wave reflecting off a surface this law must also be obeyed.

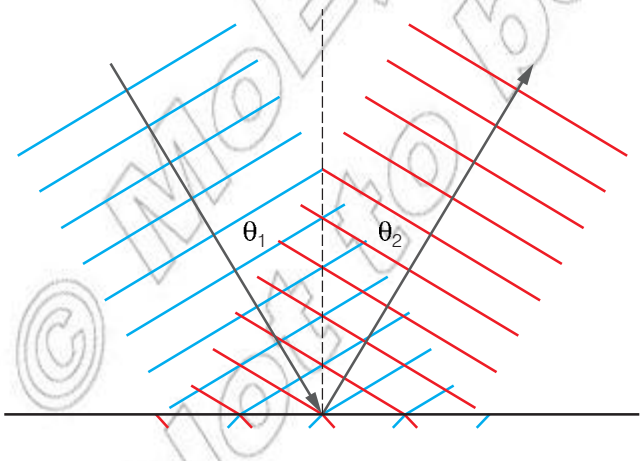


Figure 3.14 *A wave front diagram demonstrating the law of reflection*

Care must be taken to ensure this law is still valid. Notice that the wavelength of the wave does not change upon reflection.

© Not to be republished FDR Ethiopia

Activity 3.3: Wave front diagrams

Draw three wave front diagrams to show plane wave fronts reflected off a surface for the following angles of incidence.

1. 30° 2. 45° 3. 60°

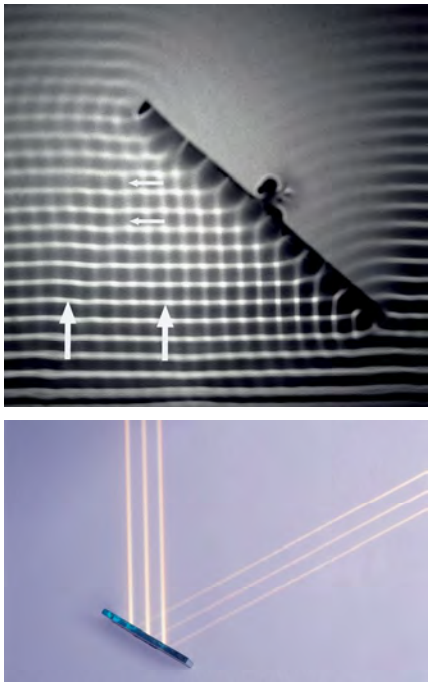


Figure 3.15 *Photos of reflection using rays of light (to show a ray diagram) and water waves (to show a wave front diagram). Notice that the law of reflection is obeyed and that there is no change in wavelength.*

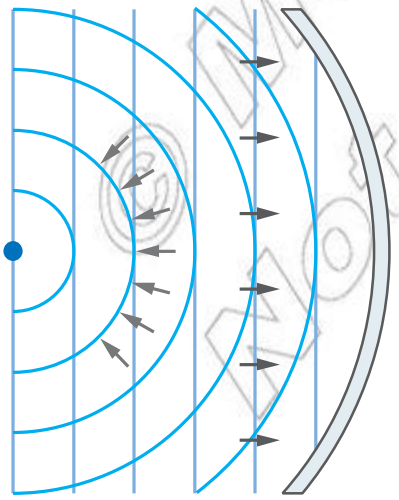


Figure 3.17 *Plane wave fronts reflecting off a circular mirror*

The diagrams below show the reflection of circular waves off a plane reflector and plane waves of a circular reflector. In both cases there is no change in wavelength.

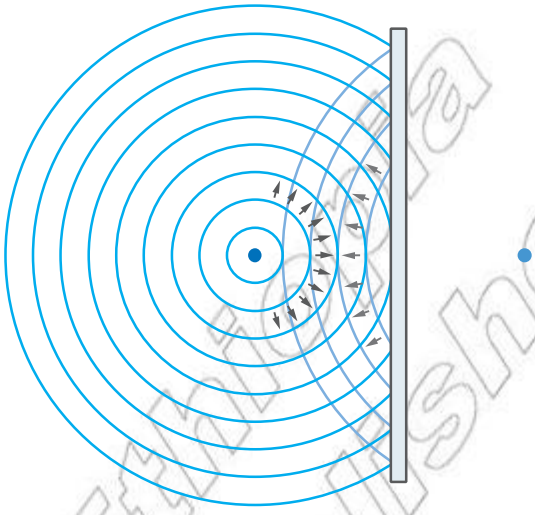


Figure 3.16 *Spherical wave fronts reflecting off a plane reflector*

Refraction in terms of wave fronts

Wave front diagrams of refraction are more complex. You recall that a wave refracts when it travels from one medium to another. As the wave enters a different medium its speed may change and so the wave bends in one particular direction.

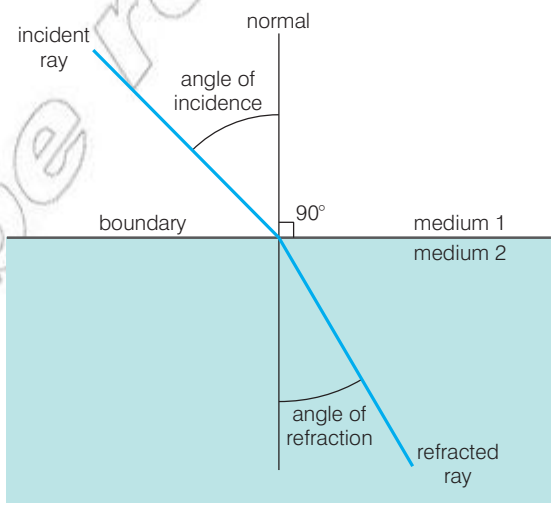


Figure 3.18 *A ray diagram to demonstrate refraction*

When a wave refracts, along with a change in speed there is also a change in wavelength.

The relationship between the angle of incidence and angle of refraction, and the wave speed in each medium is governed by Snell's law:

- $\sin \theta_1 / \sin \theta_2 = v_1 / v_2 = \lambda_1 / \lambda_2$

© Via Afrika Not to be republished

When constructing a wave front diagram of refraction, if the wave is slowing down there is a decrease in wavelength (the reverse is true if there is an increase in speed). This must be clear from the diagram.

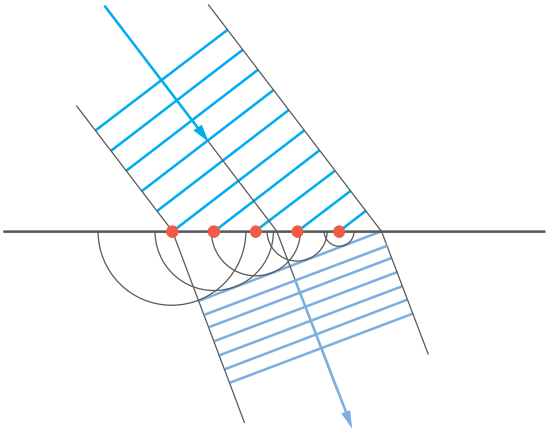


Figure 3.19 *A simple wave front diagram of refraction*

DID YOU KNOW?

The wave still obeys the wave equation $v = f\lambda$. The frequency remains constant so if the wave speed drops so must the wavelength.

Activity 3.4: Refraction

Carefully draw a wave front diagram showing a wave increasing in speed as it enters a different medium. Pay particular attention to the direction of refraction and the wavelength of the wave.

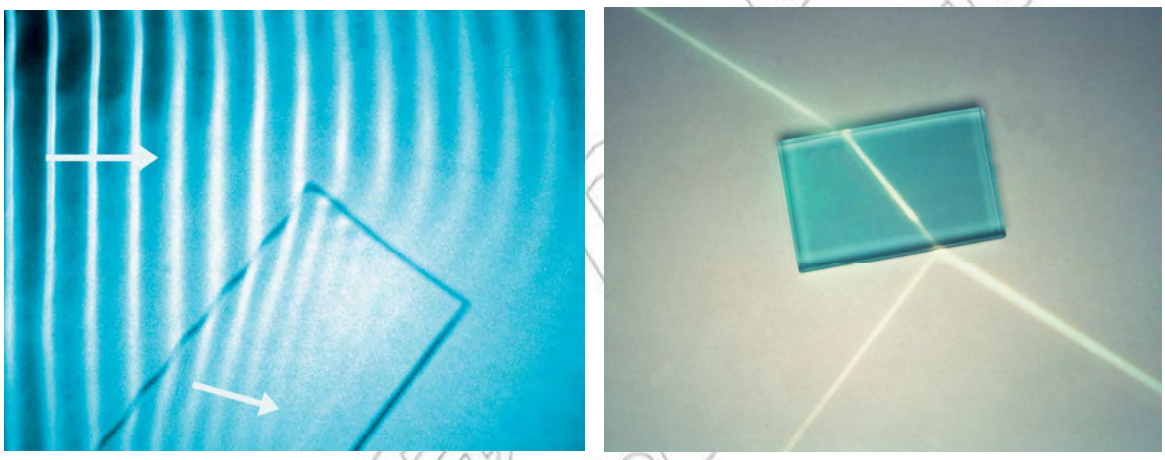


Figure 3.20 *Photos of refraction using rays of light (to show a ray diagram) and water waves (to show a wave front diagram). Notice the agreement with Snell's law: as the wave slows down its wavelength also decreases.*

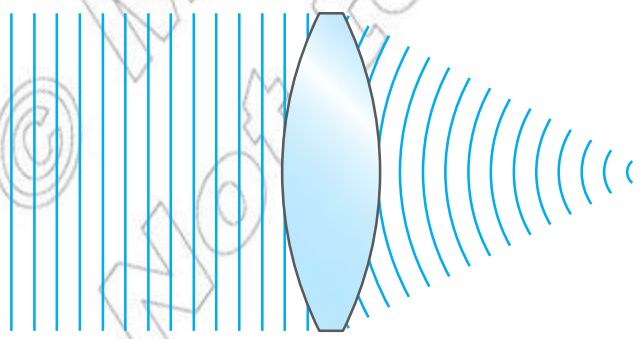


Figure 3.21 *Refracted wave fronts can also change their shape. For example, light refracted through a lens.*

Think about this...

If the wavelength of light is changing why does it look the same colour when it is inside the glass block?

© Mindset Learn Not to be republished

Summary

In this section you have learnt that:

- When constructing wave front diagrams for reflection the law of reflection must be obeyed and the wavelength of the reflected wave remains unchanged.
- When constructing wave front diagrams for refraction Snell's law must be obeyed. As a result the wavelength of the refracted wave changes.

Review questions

1. State the law of reflection and Snell's law.
2. Using Snell's law draw a wave front diagram to scale for the following refraction (including the relative size of the wavelengths):
   a) angle of incidence 60°; angle of refraction 20°
   b) angle of incidence 30°; angle of refraction 75°.

3.3 Proof of the laws of reflection and refraction using Huygens's principle

By the end of this section you should be able to:

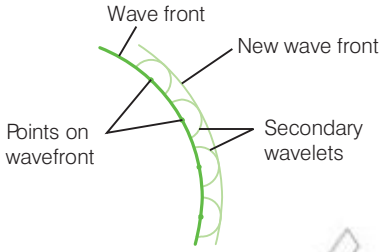
- Understand the proof of the laws of reflection and refraction using Huygens's principle.
- State the laws of reflection and refraction.
- Describe reflection and refraction in terms of the wave nature of light.

Applying Huygens's principle to reflection and refraction

Huygens used his wave theory to explain both the reflection and refraction of light.

Reflection

It is possible to confirm the law of reflection using Huygens's construction (see Figure 3.22).



The first wavefront is considered as a set of points which act as centres of disturbance. There is an infinite number of such points on the wavefront, but obviously only a finite number of them may be drawn. Each point gives rise to a new 'mini wavefront' or **secondary wavelet**, which has the same speed (and hence wavelength) as the original wavefront. The new wavefront is the line which is tangential to the secondary wavelets.

Figure 3.22 *The idea of a wavefront as a set of disturbances producing a new wave front is called Huygens' construction. Huygens' construction is an explanation for the way in which a circular wave spreads out, eventually leading to a plane wave as the radius of the circular wave becomes very large. This model of wave behaviour is useful in explaining other properties of waves.*

Refraction

Using the same technique it is possible to confirm Snell's law of refraction.

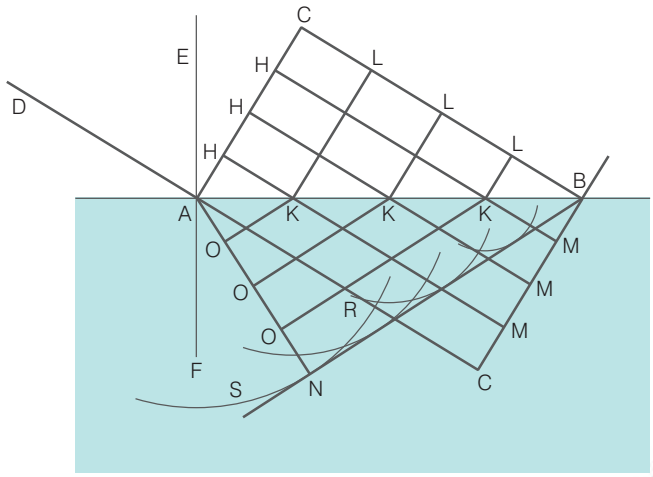


Figure 3.23 *One of Huygens's diagrams to explain refraction*

Table 3.1 *Reflection and refraction summary*

Reflection	Refraction
Law of reflection	Snell's law
angle of incidence = angle of reflection	$\sin\theta_1/\sin\theta_2 = v_1/v_2$
$i = r$ or $\theta_1 = \theta_2$	$v_1/v_2 = \lambda_1/\lambda_2$

A different model to explain reflection and refraction

At the end of the 17th century there were two competing theories concerning the nature of light. We have already encountered Huygens's wave theory. His ideas would eventually be accepted but not until the late 19th century. At the time a different theory was much more popular.

Isaac Newton proposed an alternative theory on the nature of light. He suggested that light was made up of a stream of tiny particles that he called **corpuscles** (meaning small particles). His theory took precedence over Huygens for a number of reasons.

1. Light can travel through a vacuum. No other wave motion was known to travel through a vacuum and at the time there was no theory to explain how this might be possible.
2. Light casts a sharp shadow behind opaque objects. If light was a wave diffraction would occur and the shadow would have blurry edges (we now know this does in fact happen but the wavelength of light is so small that the effect is hard to notice).

However, perhaps most important was Newton's powerful and fearsome reputation!

Newton's corpuscular theory could be used to explain all the light-related phenomena known at the time.

KEY WORDS

corpuscles *small particles believed by Isaac Newton to make up light*



Figure 3.24 *Light casts sharp shadows with no obvious diffraction.*

© Ethiopia Not to be republished





Figure 3.25 *Both Newton and Huygens proposed different theories on the nature of light.*

Reflection

This was simple to explain. Metal ball bearings thrown at a smooth steel plate bounce off the surface just like light reflects off it. Perfectly elastic particles bouncing off a surface provide a good model for the reflection of light.

Newton explained this interaction in terms of a repulsive force, which only acts near the surface of the material. When a corpuscle of light enters this region it is repelled, perfectly elastically. As his proposed force acted perpendicularly to the surface there was no change in the horizontal component of the velocity and so the corpuscle reflected off the surface at the same angle at which it approached it.

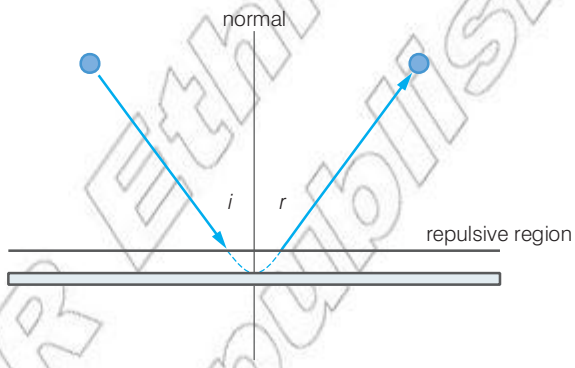


Figure 3.26 *Reflection according to the corpuscular theory of light. The angle of incidence is still equal to the angle of reflection and so the law of reflection remains intact.*

Refraction

Newton explained refraction in a similar way. If the light were to enter an optically more dense material it refracts towards the normal (e.g. air to water). He explained this in terms of a downward force acting perpendicular to the surface.



Figure 3.27 *Fresnel*

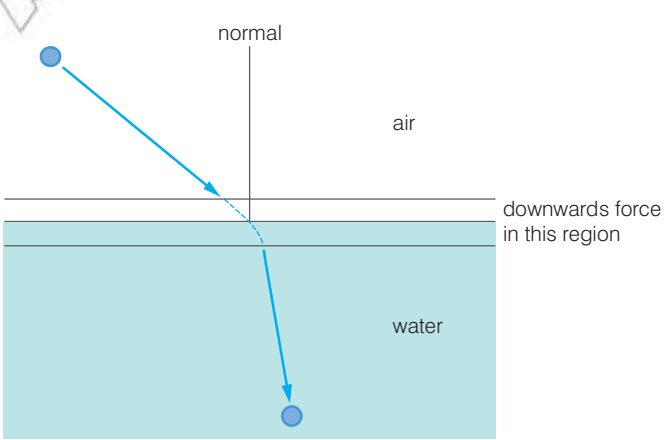


Figure 3.28 *Refraction according to the corpuscular theory of light*

FOR Ethiopia to be republished

As the corpuscle gets near the boundary between the two materials it experiences this accelerating force. As a result its vertical velocity increases, leading to a change in direction.

According to Newton light would need to travel faster in optically denser materials like glass or water.

Newton explained that the corpuscles had 'phases'. This meant sometimes the particles where repelled by the surface, and so reflect, and other times they would be accelerated towards the surface and so they were refracted. The outcome depended of the phase of the particle.

The death of the corpuscular theory

Newton's corpuscular theory remained the accepted theory for more than 100 years. In 1801 Thomas Young conducted a series of experiments on diffraction and interference that challenged the particle theory (more on this in section 3.5). However, he did not communicate his ideas in a scientific manner and so his conclusions were not widely accepted.

It was not until 1820 when the French physicist Augustin Fresnel developed a rigorous mathematical explanation of why light casts sharp shadows that the wave theory began to take precedence over Newton's ideas.

DID YOU KNOW?

Fresnel also invented a new type of lens (the Fresnel lens). The design of the lens allowed larger lenses to be constructed without a significant increase in mass and thickness. These lenses allowed more light to pass through them and so are used in applications such as lighthouses where their light may be visible over much longer distances.

The killer blow came from Jean Foucault in 1850. He was able to show that light travelled slower in optically denser materials. This was widely regarded as the conclusive piece of evidence against the corpuscular theory.

Summary

In this section you have learnt that:

- Huygens's principle may be used to prove the law of reflection and Snell's law.
- Newton's corpuscular theory of light provided an alternative explanation to Huygens's wave theory of light.

Think about this...

Modern quantum theory (specifically wave–particle duality) reintroduces the idea of particles of light: photons. However, they behave very differently to Newton's corpuscles.

Review questions

1. State the law of reflection and Snell's law of refraction
2. Demonstrate how Huygens's principle may be used to verify the law of reflection.
3. Demonstrate how Huygens's principle may be used to verify Snell's law.
4. Outline the key ideas of Newton's corpuscular theory. Include an account of the explanation for both reflection and refection.

© DBE Not to be republished

KEY WORDS

superposition *where two or more waves pass through a single point then the resulting displacement at that point is the sum of the displacements that would be created separately by each wave*

minima *areas where destructive interference of light results in a drop in intensity*

maxima *areas where constructive interference of light results in an increase in intensity*

coherent *where two waves are of the same type, have the same frequency and maintain a constant phase relationship*

3.4 Interference

By the end of this section you should be able to:

- Describe the phenomena of wave interference as it applies to light in qualitative and quantitative terms using diagrams and sketches.
- Compare the destructive and constructive interference of light with superposition along a string.
- Identify the interference pattern produced by the diffraction of light through narrow slits (single and double slits).
- Define an interferometer as a device which uses the interference of two beams of light to make precise measurements of their path difference.
- Define thin film interference and apply and use the equations to solve problems.

Diffraction of light

The phenomenon of diffraction is frequently observed with sound and longer wavelength electromagnetic waves. When these waves pass through a gap or around an obstacle they spread out.

An common example may be observed if someone is in an adjoining room and calls your name. If two rooms are connected by an open doorway the sound diffracts through the doorway and it appears that the sound comes from the doorway itself. As far as you are concerned the vibrating air in the doorway is the source of the sound itself.

With light, however, it is a different story. Unless there is a direct line of sight you will not be able to see the person who called your name. It appears that light does not diffract.

In fact this was one of Newton's key arguments against Huygens's wave theory. Even Huygens himself could not come up with a convincing counter argument.

We now understand that the amount of diffraction depends on the wavelength of the wave relative to the size of the gap. In simple terms, the closer the size of the gap is to the wavelength the better or more pronounced the diffraction.

Light has a very short wavelength and so a gap as large as an open doorway produces very little diffraction. In order to observe diffraction of light a much smaller gap is needed.

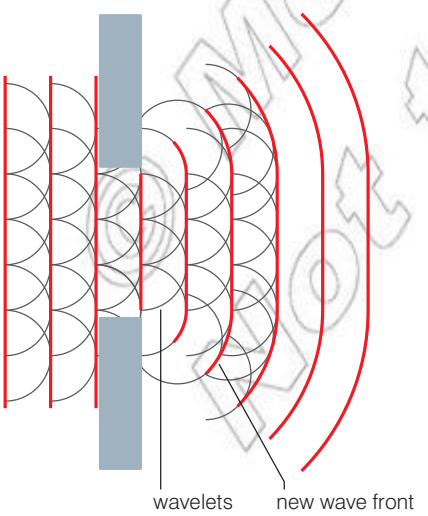


Figure 3.29 *Using Huygens's principle to explain diffraction*

Diffraction and Huygens's principle

Huygens's principle may be used to explain the phenomena of wave diffraction.

When light passes through a small gap every point of the light wave within the gap creates its own circular wavelet. The gap therefore

© MoE Free Textbook Not to be Republished

effectively creates a new wave source. These wavelets travel out and form a new wave front. This may be seen in the diagram in Figure 3.29.

The centre of this new wave front has the highest intensity, with the intensity falling towards the edge. At the end of this new wave front another circular wavelet is created; this leads to the edges of the wave front bending around, as shown in the diagram. The result of this effect is the new wave front and so the wave itself spreads out.

Interference of light and interference patterns

The principle of **superposition** was discussed in the previous unit. This principle also applies to light. If light waves are made to superpose then the intensity of the light at that point may increase or decrease.

Destructive interference gives rise to a drop in intensity, or dark patches (called **minima**). Constructive interference results in an increase in intensity, or brighter regions (called **maxima**)

This effect is similar to that observed by the superposition of the waves travelling along a string. You may recall where the waves along the sting are in antiphase they cancel out, giving rise to a node. This is the equivalent of a minima. Here the intensity falls to zero, just like the amplitude of the oscillations on the string.

The antinodes on the string are created when the waves are in phase, giving constructive interference and so maximum amplitudes of oscillations. This is the same at a maxima of light; this is where the intensity of the light is at its greatest.

If light is made to superpose, an interference pattern may be formed. This is a series of maxima and minima. These bright and dark patches may be observed on a screen.

In order to create a sustained interference pattern two **coherent** sources of light must be used (more on this in section 3.6). In order to create a source of light a slit is often used. This diffracts the light as it passes through it and so the slit acts as a source of light. Several slits are used to acts as several sources of light.

We shall look at the formation of two different interference patterns, one created by diffracting light through a pair of slits (a double slit) the other by diffracting light through a single slit.

Double slit

Using a simple double slit, an interference pattern like the one in Figure 3.30 may be observed.

This series of maxima and minima is created as the light diffracted from each slit superposes.

As the light passes through slit A it diffracts and so spreads out. The same effect occurs at slit B. We have effectively produced two sources of light.

Activity 3.5: Diffraction

Carefully copy the diagram above to show how Huygens's principle may be used to explain diffraction. For gaps equal to the wavelength of the wave this effect is even more pronounced and so the wave spreads out even more. Using diagrams can you show why?

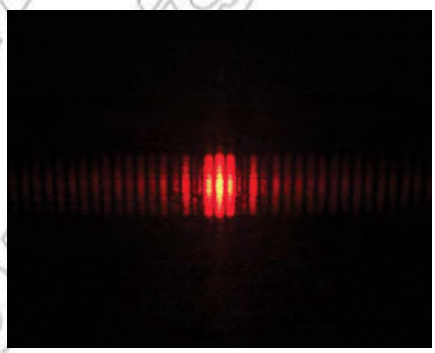


Figure 3.30 *The interference pattern produced by a double slit*

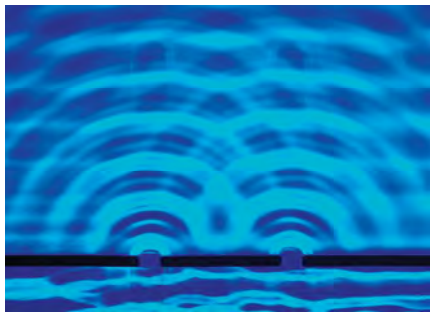


Figure 3.31 *The same effect may be more easily observed using a ripple tank. Here two vibrating sources create an interference pattern.*

© MOED-EPDC Ethiopia
Not to be reproduced

KEY WORDS

fringes *light or dark bands produced by the diffraction or interference of light*

As the light diffracts from each slit it overlaps and superposition occurs. This produces a noticeable interference pattern on the screen.

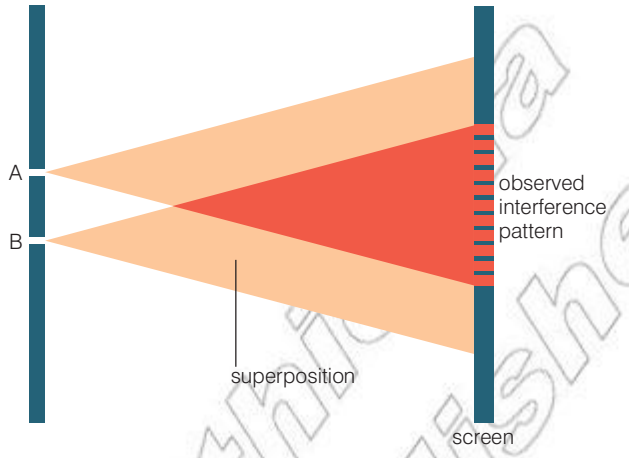


Figure 3.32 *The formation of interference pattern produced by a double slit*

Looking carefully at the interference pattern you can see that it is a series of bright and dark **fringes** of equal width. The brightest fringe is located in the middle and is called the central maximum (or occasionally the zero-order maximum). The bright fringes either side are called the first-order maxima, followed by the second-order maxima, etc.

A simple sketch of intensity against distance may be seen in Figure 3.33:

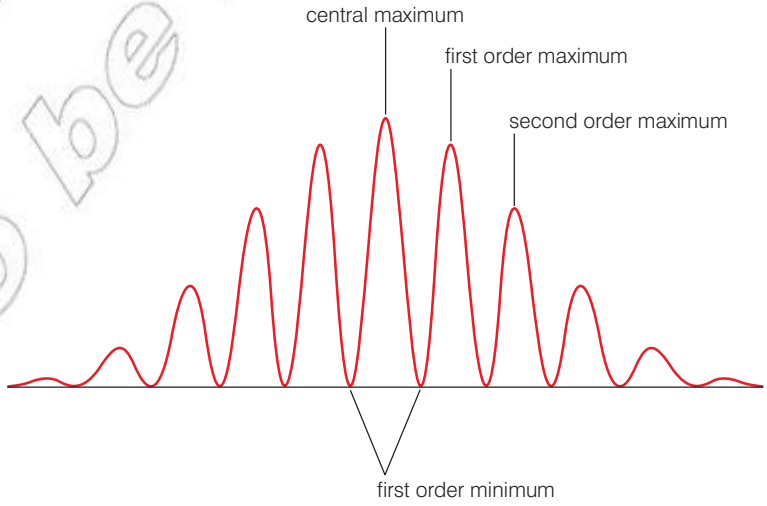


Figure 3.33 *The intensity varies through a series of maxima and minima. The greatest intensity occurs at the central maximum.*

These alternating maxima and minima are formed due to the light from each slit interfering. At the central maximum, the light striking the screen from slit A has travelled the same distance as the light from slit B. As a result, the waves are in phase (assuming the light at A and B is in phase), and so constructive interference occurs.

© MoE, PDK Ethiopia
Not to be republished

However, at the first-order minima the light from each slit has had to travel a different distance. This is referred to as the **path difference**. The light from one slit travels further and, at the minima, arrives in anti-phase with the light from the other slit. There is a **phase difference** of π and so destructive interference occurs. The light from one slit has travelled exactly half a wavelength further and so a peak meets a trough.

This process continues as you move along the screen creating a series of maxima and minima.

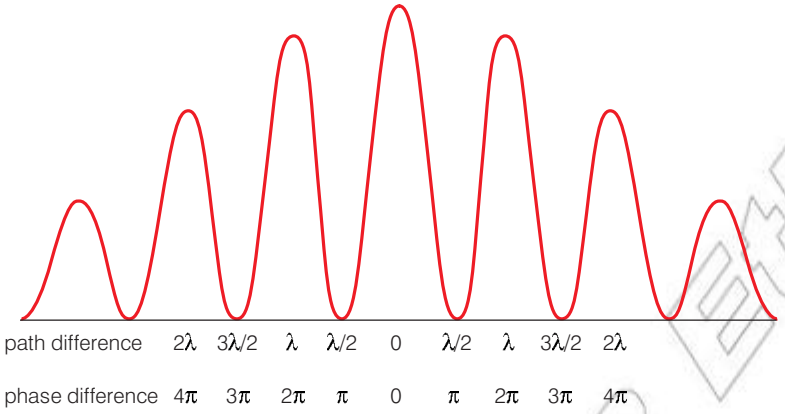


Figure 3.34 *The path differences and subsequent phase differences of the light from each slit give rise to a sustained interference pattern.*

There is a more detailed mathematical treatment of this effect in section 3.5. However, in general:

- **For constructive superposition, path difference = $n\lambda$**
- **For destructive superposition, path difference = $(n + \frac{1}{2})\lambda$**

where n = 0, 1, 2, 3, etc, depending on the maxima/minima.

Single slit

A different interference pattern is observed when light passes through a single slit. This may be seen in Figure 3.36 with relative intensities shown in Figure 3.37.

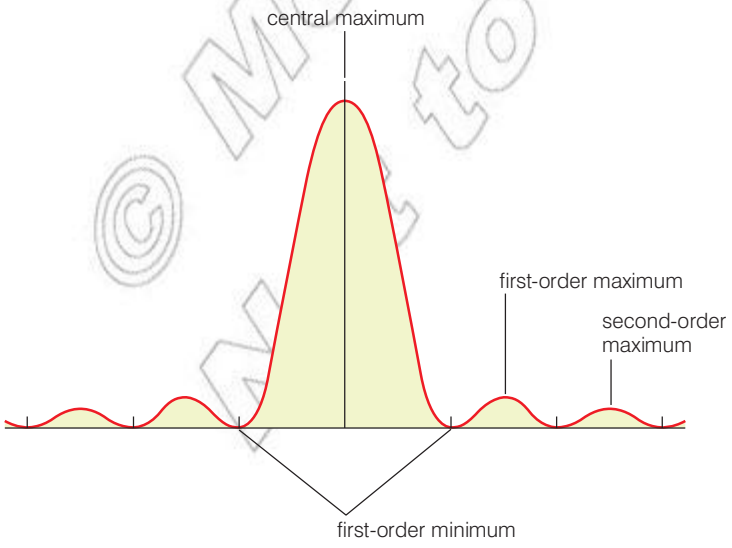


Figure 3.37 *The relative intensity of the maxima produced by single slit diffraction*

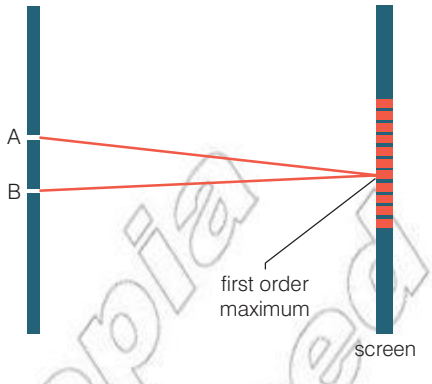


Figure 3.35 *At the first-order maxima the path difference is exactly one wavelength. This means that the light from each source is back in phase and so constructive interference occurs and maxima are observed.*

KEY WORDS

path difference *the difference in distance travelled by light diffracted by separate slits*

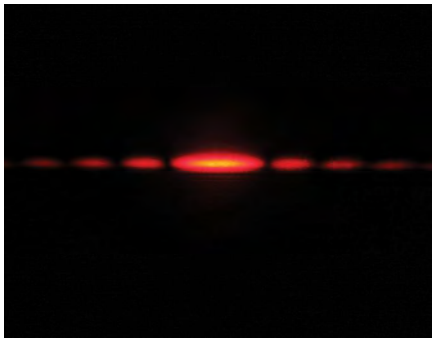


Figure 3.36 *The interference pattern produced by a single slit*

© MOE, EDP Ethiopia Not to be republished



Figure 3.38 *Again a ripple tank may be used to help see this effect. A wide central maximum is observed with minima either side.*

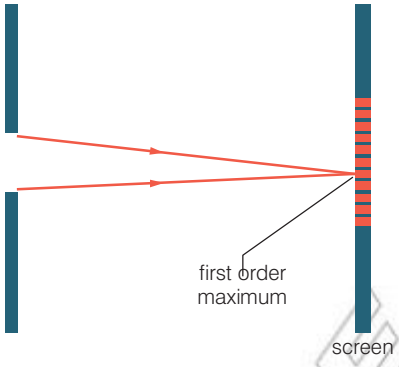


Figure 3.40 *In this diagram the size of the slit has been exaggerated to demonstrate the path difference between the light from the top and bottom of the slit.*

Again there is a series of bright and dark fringes. However, this time the central maximum is twice as wide as the first-order maxima and much brighter.



Figure 3.39 *The central maximum observed in the interference pattern created by a single slit is twice as wide as the first-order maxima (a = 2b).*

This interference pattern is created as the light diffracted from the extremes of the slit has had to travel different distances.

There is a path difference, which leads to a phase difference between the light from the top and bottom of the slit. This produces either constructive or destructive interference. There is a more detailed mathematical treatment of this effect in section 3.7.

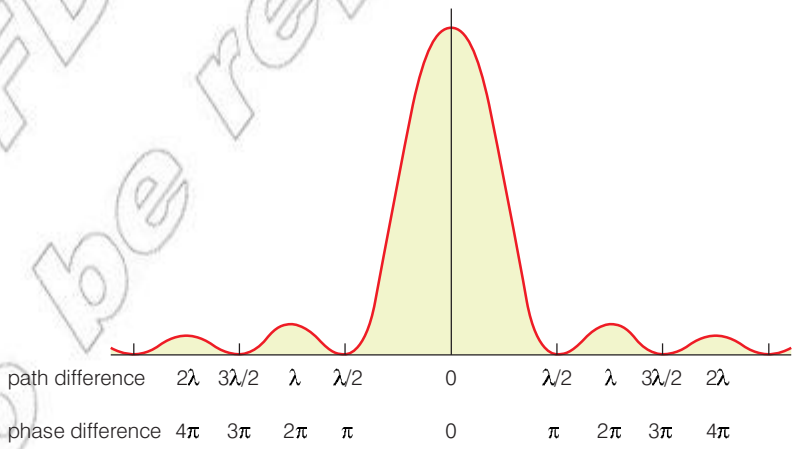


Figure 3.41 *The path differences and subsequent phase differences of the light from each part of the single slit gives rise to a sustained interference pattern.*

The interferometer

An interferometer is a simple optical device that makes uses of the interference of light to determine the wavelength of the light.

There are several different designs but most involve splitting a beam of light into two different beams. Each beam then travels a carefully controlled distance before reflecting back.

The two waves then interfere and the resulting interference pattern may be used to determine the wavelength of the wave (or

© KloE, FDE Etwopia Not to be republished

occasionally some properties of the medium or material it reflects off of).

A common example is the **Michelson interferometer**.

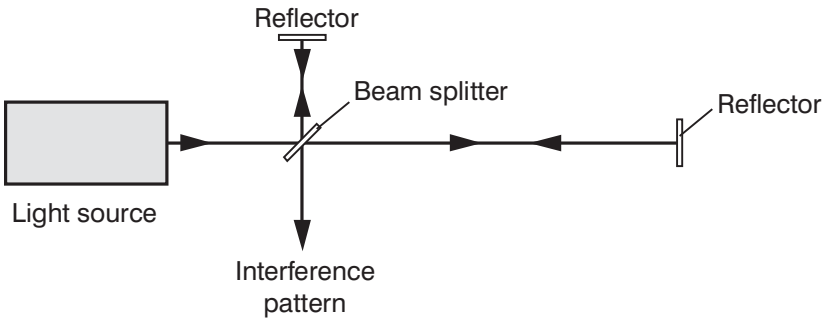


Figure 3.42 *The Michelson interferometer was invented by Albert Abraham Michelson. It employs a single beam splitter for separating and then recombining the beams of light.*

The two beams (the horizontal and the vertical) travel different distances. This creates a path difference. There is therefore constructive or destructive interference in the output beam (the one at the bottom). By carefully varying this path difference the wavelength of the light may be determined.

KEY WORDS

Michelson interferometer *a device which uses the property of interference to determine the wavelength of light thin-film*

thin-film interference *colourful patterns caused by the interference of light waves reflecting off the top surface of a thin film with those that have reflected off the bottom surface*

Thin-film interference

Constructive and destructive interference of light waves is also the reason why we see colourful pattern in soap bubbles or on the surface of a puddle of oil.

This effect is known as **thin-film interference**. It is due to the interference of light waves reflecting off the top surface of a film with those that have reflected of the bottom surface of the film. The effect is only colourful if the film is very thin, close to that of the wavelength of light.

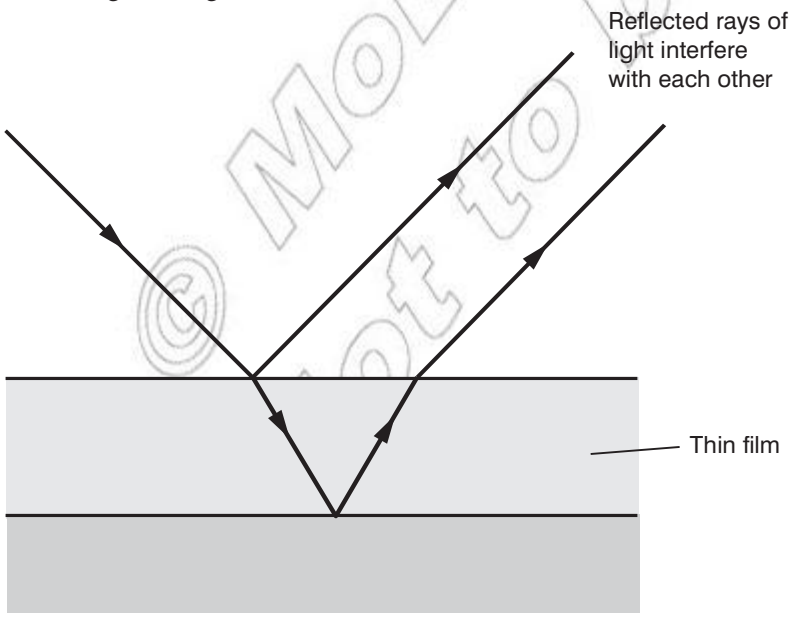


Figure 3.43 *Thin-film interference*



Figure 3.44 *Thin-film interference leads to colourful patterns*

© MoE Fix Ethiopia Not to be republished

Very importantly if the wave reflects off a surface of lower refractive index than the medium it is travelling through then there is no phase change in the reflected wave. However, if the reflecting surface has higher refracted index there is a 180° phase shift in the reflected wave. This is equivalent to a path difference of half a wavelength.

Therefore light travelling through air will undergo a 180° phase shift when it reflects off almost any surface (water, oil, glass, etc). All of these materials have a higher refractive index than air.

In order to get constructive interference the two reflected waves must have a path difference equal to an integral number of wavelengths (1, 2, 3, etc). However, we must also take into account any phase change caused by the light reflecting off either surface where the refractive index is higher.

For example, if the thin film is oil floating on water you get constructive interference if the oil is 1/4, 3/4, 5/4, etc. of the wavelength of light. The light reflecting off the top surface undergoes a phase change of 180°. The light travelling through the oil reflects off the bottom surface; however, as oil has a higher reflective index than water then there is no phase change. This reflected ray travels back up to the surface. It has travelled a distance equal to $2 \times \frac{1}{4}\lambda = \frac{1}{2}\lambda$.

This wave is now in phase with the wave reflected from the surface (having undergone a 180° phase change on reflection), and so constructive interference occurs.

As the oil moves around tiny changes in thickness leads to constructive or destructive interference of different colours of light. The different colours have different wavelengths and so in order for constructive interference the thickness must be exactly $\frac{1}{4}\lambda$. This gives rise to the coloured pattern you see.

The formula used for thin-film interference is:

- $path\ difference_{max} = 2t + \Phi$

where

t = the thickness of the thin film

Φ = the net phase change between the two reflected rays (top surface and bottom surface) expressed as a path difference*

path difference$_{max}$ = the maximum path difference.

* If both reflections occur at boundaries with a material of lower refractive indices then neither wave is inverted and their net phase difference is zero. If both reflections occur at boundaries with a material with higher refractive indices then both waves are inverted and their net phase difference is zero. However, if one reflection occurs at a boundary with a higher refractive index and the other at a lower refractive index (or vice-versa) then there is a net phase difference between the reflections of 180° and this equates to $\frac{1}{2}\lambda$.

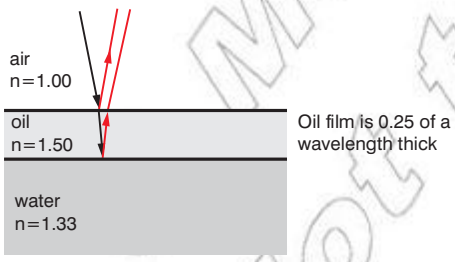


Figure 3.45 *Thin-film interference from oil on the surface of water.*

MOE-EDUTRON is not to be republished

For constructive interference the maximum path difference must be equal to an integer number of complete wavelengths (e.g. $m\lambda$ where $m = 0, 1, 2$, etc). For destructive interference the maximum path difference must be equal $(m + ½)\lambda$ (e.g. ½λ, 3/2λ, etc.).

So for constructive interference: $(m + ½)\lambda = 2nt$

For destructive interference: $m\lambda = 2nt$

Where n is the refractive index of the film.

Summary

In this section you have learnt that:

- Light diffracts when it passes through a gap similar in dimensions to the wavelength of the wave.
- Two sources of coherent light may superpose and form an interference pattern.
- An interference pattern is the result of a path difference and so a phase difference between different rays of light.
- For constructive superposition, path difference = $n\lambda$, and for destructive superposition, path difference = $(n + ½)\lambda$.
- The interference pattern produced by a double slit comprises of a series of equal width maxima and minima known as fringes.
- The interference pattern produced by a single slit comprises a wide central maximum (twice the width of subsequent maxima) with minima either side.
- Thin film interference is due to the interference of light waves refecting off the top surface of a film with those that have reflected off the bottom surface of the film.

Review questions

1. Describe how Huygens's principle may be used to explain the phenomenon of diffraction.
2. Explain how an interference pattern may be formed from two coherent sources of light.
3. Explain the meanings of the terms path difference and phase difference and relate them to the interference pattern produced by a double slit.
4. Describe the similarities and differences between the interference pattern produced by a double slit and the pattern produced by a single slit.

© Ethiopia Not to be published

3.5 Young's double slit experiment and expression for fringe width

By the end of this section you should be able to:

- Explain the interference of Young's double slit experiment.
- Carry out calculations involving Young's double slit experiment.

Young's double slit experiment

The interference effects of light were first demonstrated by Thomas Young back in the early part of the 19th century.

He used two narrow slits to produce an interference pattern from a light source, as described in section 3.4. However, before the light passed through the slits Young also used a single monochromatic filter and a single slit (more on the reasons for this in section 3.6).

A diagram of Young's experiment may be seen in Figure 3.46.

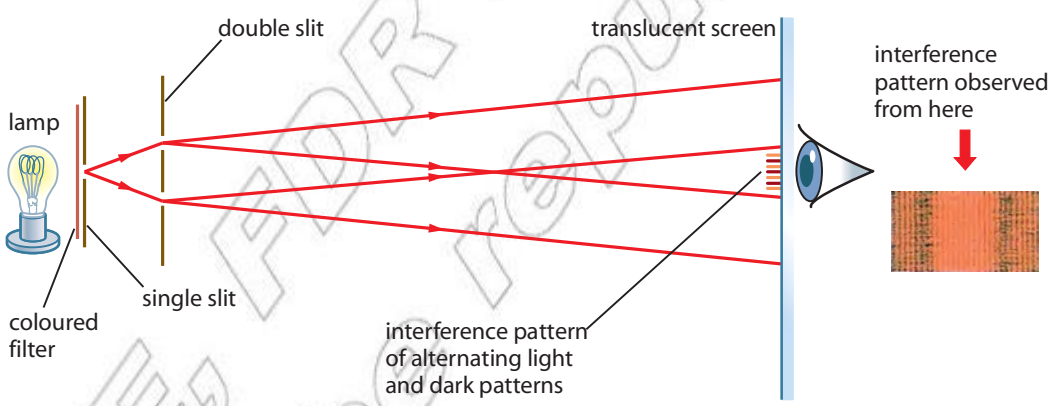


Figure 3.46 *Young's double slit experiment*

Explaining interference

The production of an interference pattern by two parallel slits can be explained by thinking about the phase of the waves arriving at the screen. Figure 3.47 exaggerates the distance between the two slits in order to make the situation clearer, and explains how the pattern of light and dark bands (often referred to as fringes) arises. A light band will be a region of constructive interference, as waves superimpose in phase, so the *difference* in distance travelled to the screen by waves from A and B must be a whole number of wavelengths $n\lambda$. A dark band will be a region of destructive interference, as waves superimpose antiphase, so the difference in distance travelled by the waves will be a number of half wavelengths $(n\lambda/2)$.

© MOE, FDR Ethiopia Not to be republished

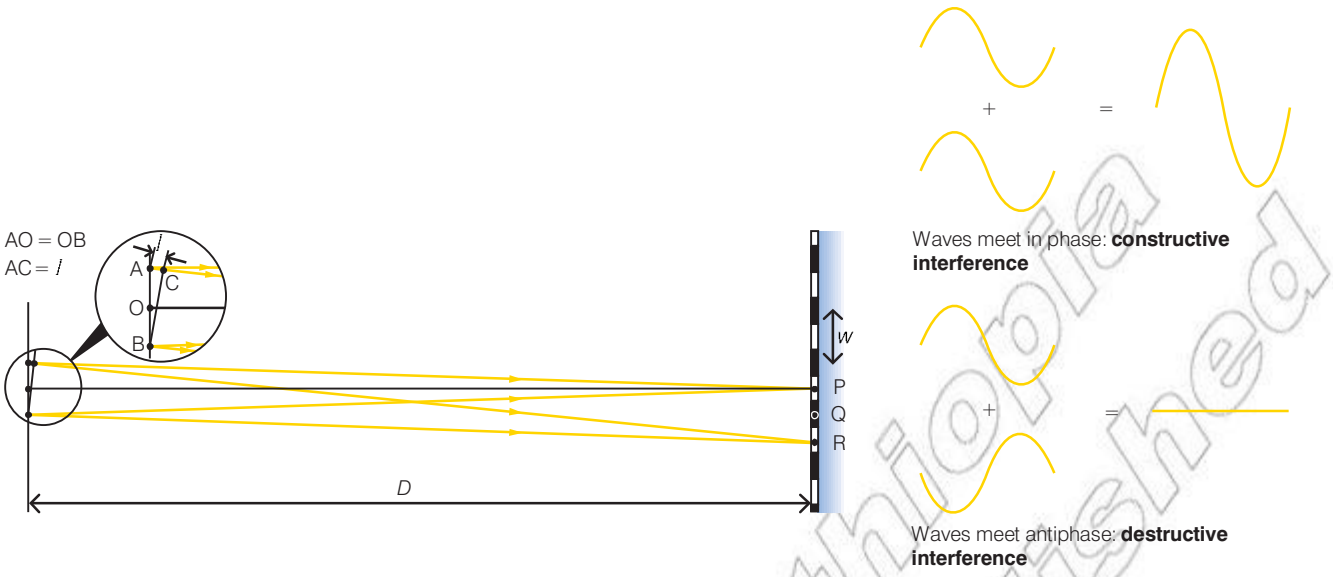


Figure 3.47 *Light waves leave slits A and B in phase. Since AP = BP, the waves must arrive at P in phase, so constructive interference occurs here and a bright area is seen. The distance AR is exactly one wavelength more than the distance BR, so the waves also arrive at R in phase, leading to a bright area here also. The distance AQ is exactly half a wavelength more than the distance BQ, so the waves arrive at Q antiphase, resulting in a dark area.*

Some reasonably straightforward trigonometry enables us to derive a relationship between D the distance from the slits to the screen, w the distance between successive light or dark fringes, s the slit separation and λ the wavelength of the light used. Referring again to Figure 3.47, we know that:

$AR - BR = \lambda$

If angle ABC = θ, then:

$$\sin\theta = \frac{AC}{AB} = \frac{\lambda}{s}$$

But the diagram also shows that $\sin\theta = PR/OR$. Since θ is very small (remember that this diagram exaggerates the position – in practice PR is about 2 mm while OP is about 1 m) OR ≈ OP, so we can write:

$$\sin\theta \approx \frac{PR}{OP}$$

Since PR = w and OP = D, we can use the expression $\sin\theta = \lambda/s$ given above and write:

$$\frac{\lambda}{s} = \frac{w}{D} \quad or \quad \lambda = \frac{ws}{D}$$

Phase difference and path difference

The difference between the distance travelled by one ray and another is called the path difference, Δx. This path difference causes a phase difference for the two rays. These two are related by the equation

$$\text{phase difference} = \frac{2\pi\Delta x}{\lambda}$$

where λ is the wavelength of the light. The phase difference is in radians (2π radians = 360°).

© MoE, FDR Ethiopia Not to be republished

Worked example 3.1

Light with an unknown wavelength passes through two narrow slits 0.3 mm apart and forms an interference pattern on a screen 2.0 m away from the slits. If the distance between the fringes in the interference pattern is 3 mm, what is the wavelength of the light?

We know that the wavelength λ, the fringe separation w, the slit separation s and the distance D from the slits to the screen are related by the equation:

$$\lambda = \frac{ws}{D}$$

so we may substitute the values known into the relationship:

$$\lambda = \frac{3 \times 10^{-3}\ \text{m} \times 0.3 \times 10^{-3}\ \text{m}}{2.0\ \text{m}} = \frac{0.9 \times 10^{-6}\ \text{m}^2}{2.0\ \text{m}}$$

$$= 4.5 \times 10^{-7}\ \text{m}$$

The wavelength of the light is 4.5×10^{-7} m, at the violet end of the spectrum. Wavelengths of light are often expressed in nm (nanometres). 1 nm = 10^{-9} m, so this light has a wavelength of 450 nm.

Summary

In this section you have learnt that:

- Young's double slit experiment provides evidence for the wave nature of light.
- For interference from a double slit $\lambda/s = w/D$.

Review questions

1. Explain how Young's experiment produced an interference pattern.
2. With the aid of diagrams, show how $\lambda/s = w/D$.
3. Explain the effect on the interference pattern (fringe width) of:
   a) using light with a higher frequency
   b) using narrower slits
   c) increasing the distance between slits and screen.
4. A laser produces an interference pattern on a screen 10 m from a pair of slits. The slit space is equal to 0.25 mm and the fringe width is measured to be 26 mm. Determine the wavelength and frequency of the light from the laser.

© MoE-DVV Not to be copied

3.6 Coherent sources and sustained interference of light

By the end of this section you should be able to:

- State the conditions necessary for the interference of light to be shown.

What does coherent mean?

Remember if waves are not coherent they will not produce a stable interference pattern. A stable interference pattern is often referred to as sustained interference of light.

In order for two waves to be coherent they must:

- be the same type of wave

  It is not possible to produce a stable interference pattern with an electromagnetic and a sound wave!

- have the same frequency

  Otherwise 'beats' will occur. If both waves are the same frequency, it therefore follows they are travelling at the same speed and have the same wavelength as the waves are in the same medium.

- maintain a constant phase relationship.

  The waves do not have to be in phase; however, their phase relationship must be constant (0, π, 2π, etc.). This ensures that at any given distance the type of interference is always the same (i.e. constructive or destructive). If the phase difference was changing – for example, one source moving relative to the other, or one source starting and then stopping – then the interference observed at any position would be change. As a result a stable interference pattern would not form.

The degree of coherence is often measured by the interference visibility. This is simply a measure of how perfectly the waves cancel out due to destructive interference.

How did Young make sure his light was coherent?

Looking back at Young's experiment he included a monochromatic filter and a single slit. Both of these are necessary to ensure the light is coherent.

1. Why the filter as well as the double slit?

Figure 3.46 on page 128 shows a filter between the light source and the double slit. This is necessary to ensure that the light used in the experiment is monochromatic or of one wavelength only (although in fact filters usually allow through a range of wavelengths rather than a single wavelength). Without the filter the fringes are blurred and consist of a range of colours:

© Ministry of Education Ethiopia Not to be republished

$$\lambda = \frac{ws}{D} \text{ or } w = \frac{\lambda D}{s}$$

Thus if there is a range of wavelengths, there will be a range of fringe separations too. A filter is not necessary if a monochromatic source is used. A sodium lamp is effectively monochromatic, since the intensity of the light emitted by it at a wavelength of 5.89×10^{-7} m is many times that emitted at other wavelengths. Many lasers also produce monochromatic light.

2. Why the single slit as well as the double slit?

Even if a monochromatic source is used, the light emitted from it contains many imperfections. The source emits light due to the loss of energy by excited electrons within the atoms of the source. Different parts of the source therefore emit light at slightly different times and with different phases. Although this incoherence happens so rapidly that it is invisible to our eyes, it makes interference between light from two different parts of a source impossible to observe. The single slit in front of the source therefore ensures that the light reaches both slits in phase, so that the slits act as sources of waves rather like the dippers used to produce two simultaneous circular waves in the ripple tank in Figure 3.31. Interference could not be observed if the dippers did not have a constant phase relationship. (The use of laser light overcomes these problems, since laser light is coherent – there is a constant phase relationship between all parts of the source.)

Using a laser

A modern day version of Young's experiment involves the use of a laser. Here it is not necessary to use a single slit, nor a monochromatic filter. The light from the laser is already coherent; all that is required is that the light must pass through a double slit. Each slit acts as a source of light and a stable interference pattern is observed.

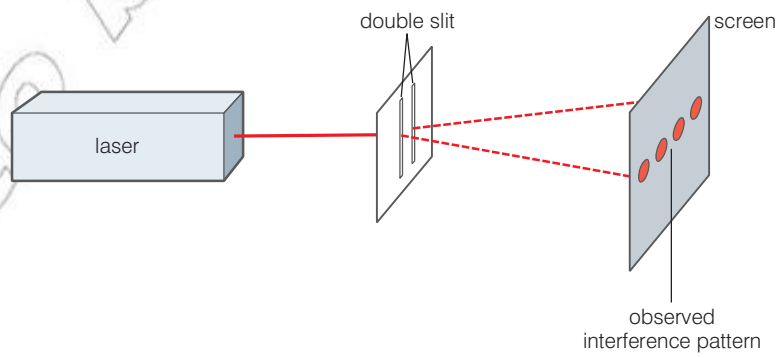


Figure 3.48 *Using a laser to produce a sustained interference pattern*

© MoE. Not to be republished

Summary

In this section you have learnt that:

- In order to form a sustained interference pattern the light must be coherent.
- Coherent waves have the same frequency and a constant phase relationship.
- Lasers produce coherent waves when passed through a double slit.

Review questions

1. Explain the meaning of the term coherent.
2. Explain why an interference pattern is not observed between the light produced from a pair of car headlights.

3.7 Diffraction due to a single slit and a diffraction grating

By the end of this section you should be able to:

- Describe the diffraction due to a single slit, including the interference caused rays of light coming from different parts of the slit.
- Describe and explain the diffraction of light in quantitative terms using diagrams.
- Describe the effects of using a diffraction grating.

What causes single slit diffraction?

We have already outlined this effect in section 3.4. This section provides a more mathematical treatment of the effect.

Just as for refraction and reflection, the behaviour of light as it passes through a narrow slit like this can be explained using Huygens' construction. Figure 3.49 shows how this is done.

Note that we assume that waves reaching the slit are plane waves travelling in a direction perpendicular to the slit. As the wave passes through the slit, each point on the wave may be considered to act as the source of a new, circular wavefront, as we saw earlier. This means that a plane wavefront with the same width as the slit will travel away from the slit in the same direction as the original wave was travelling. Now consider a direction that makes an angle θ with the original direction of travel in such a way that there is a path difference of one wavelength

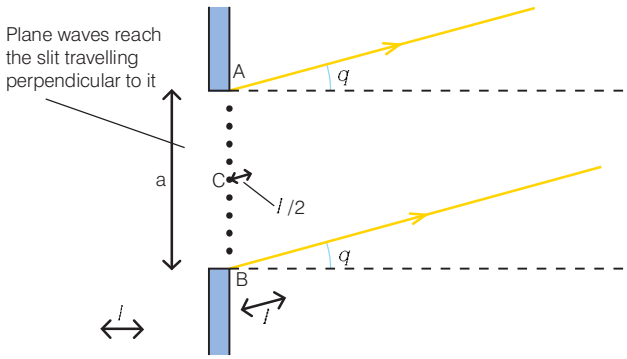


Figure 3.49

© MoE: Ethiopia Not to be republished

between the wavelet from A and that from B (see Figure 3.49). Point C is midway between points A and B. The wavelet from point C is therefore exactly antiphase with the wavelet from point A, and so the two wavelets can cancel out. For every secondary wavelet formed at a point along AC, there can always be found another secondary wavelet from a point along BC with which the wavelet can cancel. In this way all the light coming from AC cancels out all the light coming from BC, no light energy flows at angle θ to the original direction of travel, and a dark band appears on the screen in this direction. From the diagram it can be seen that:

$\sin \theta = \lambda/a$

Since the conditions for light from the two halves of the slit cancelling each other will also exist when the path difference is 2λ, 3λ, and so on, it follows that in general the minima of intensity occur at angles given by:

$\sin \lambda = n\lambda/a$, where $n = 1, 2, 3 \ldots$

At points between these angles not all the light is cancelled and so light bands appear. The intensity of these bands decreases as θ increases since more and more secondary wavelets are available to be paired with others out of phase as the angle gets bigger. (Note that this analysis is strictly only true if the waves reaching the screen are still plane waves – this will only be so if the distance from the slit to the screen is very large compared to the width of the slit.)

For small angles, $\sin \theta \approx \theta$ if θ is in radians, and so we can write $\theta = n\lambda/a$. If light of wavelength 6×10^{-7} m passes through a slit 3×10^{-4} m wide, the angle between the centre of the pattern and the first minimum will be given by:

$$\theta = \frac{6 \times 10^{-7}\ \text{m}}{3 \times 10^{-4}\ \text{m}} = 2 \times 10^{-3}\ \text{radians}$$

If the distance from the slit to the screen is 1 m, this represents a distance on the screen of:

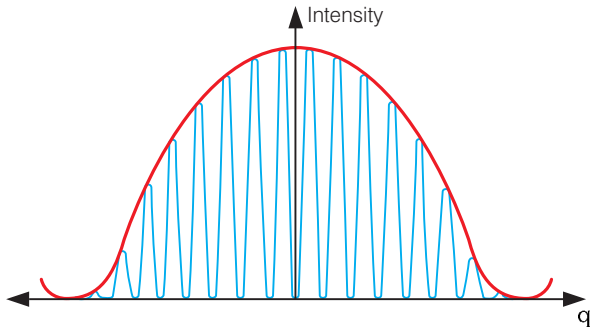
2×10^{-3} radians × 1 m = 2×10^{-3} m = 2 mm

In examining the two-slit experiment earlier, we made the (unstated) assumption that the width of each slit was small compared with the wavelength of the light ($a < \lambda$), so that the diffraction pattern produced by each slit was very wide. (Because $\sin \theta = n\lambda/a$, the angle between the centre of the pattern and the first minimum is given by $\sin \theta = \lambda/a$. If $a < \lambda$, $\lambda/a > 1$ and so the central peak of the pattern is so wide that it effectively covers all angles.) This meant that each slit effectively produced an even light intensity over a wide angle, and the pattern of light intensity produced by the two slits was entirely due to interference between them, producing the even light and dark bands of Figure 3.46.

If the width of the slits is not narrow compared with the wavelength of the light, each slit produces its own Fraunhofer diffraction pattern, and these then interfere to produce an overall pattern. The way in which the distribution of light in this overall pattern is determined is shown in Figure 3.50 – the graph of intensity versus

© MoE, Ethiopia
Not to be republished

angle is actually the product of a diffraction curve (describing Fraunhofer diffraction at each slit) and an interference curve (describing the superposition of light from each slit). If I_s is the intensity at a point due to interference and I_d is the intensity at the same point due to diffraction, the resultant intensity I is given by $I = I_s \times I_d$. If I_d is zero at any point, then $I = 0$ at that point regardless of the value of I_s.



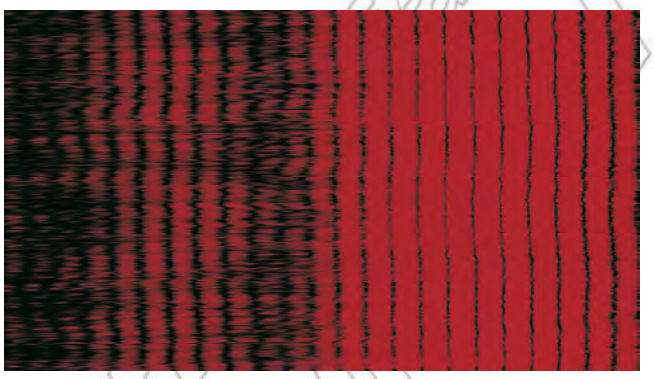


Figure 3.50 *As a result of Fraunhofer diffraction at each slit, the overall intensity distribution produced by a pair of slits which are wide compared with the wavelength of light looks like this. The photograph shows how the pattern consists of a complex series of bands whose intensity varies widely, rather than the series of light and dark bands produced by a narrow pair of slits.*

Issues caused by diffraction in optics

As light diffracts whenever it passes through an aperture it can cause some unwanted effects resulting in blurring or unclear images.

As we've already seen if light is diffracted through a circular aperture rather than a single slit a slightly different interference pattern is observed.

The pattern consists of a central bright spot called the Airy disc surrounded by concentric light and dark rings. These rings are the result of constructive and destructive interference, respectively. The bright rings are much fainter than the Airy disc and just like the maxima studied in the previous sections their intensity decreases with distance from the centre.

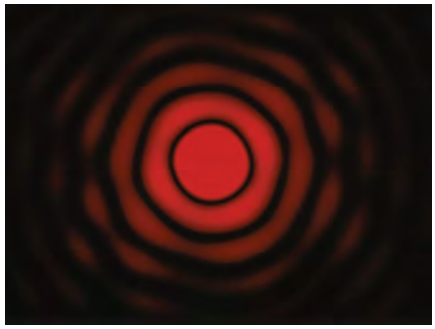


Figure 3.51 *The interference pattern created when light diffracts through a circular aperture rather than a single slit.*

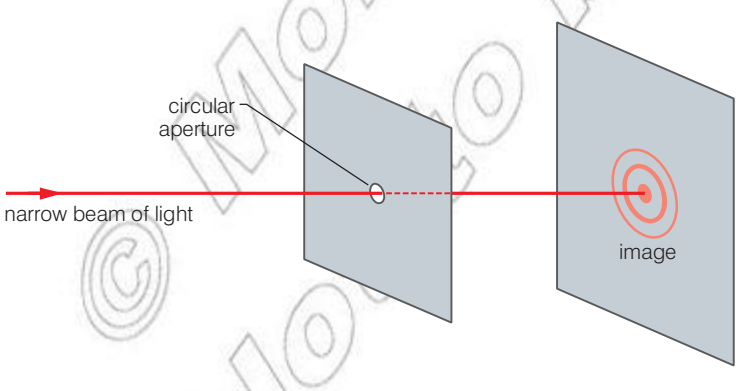


Figure 3.52 *The interference pattern from a circular aperture*

This diffraction creates a problem when observing two different light sources, for example, light from distant stars or light reflected from two different parts of an object under a microscope.

As the light passes through the aperture of the telescope or microscope it diffracts. This creates two (or more) interference patterns like the one seen above.

© MoE FET Not to be reproduced

If the Airy discs created by each ray do not overlap significantly it is quite easy to resolve two images. However, if the Airy disc created by one ray lies inside the first-order maxima of the second then it becomes very hard to see two distinct images.

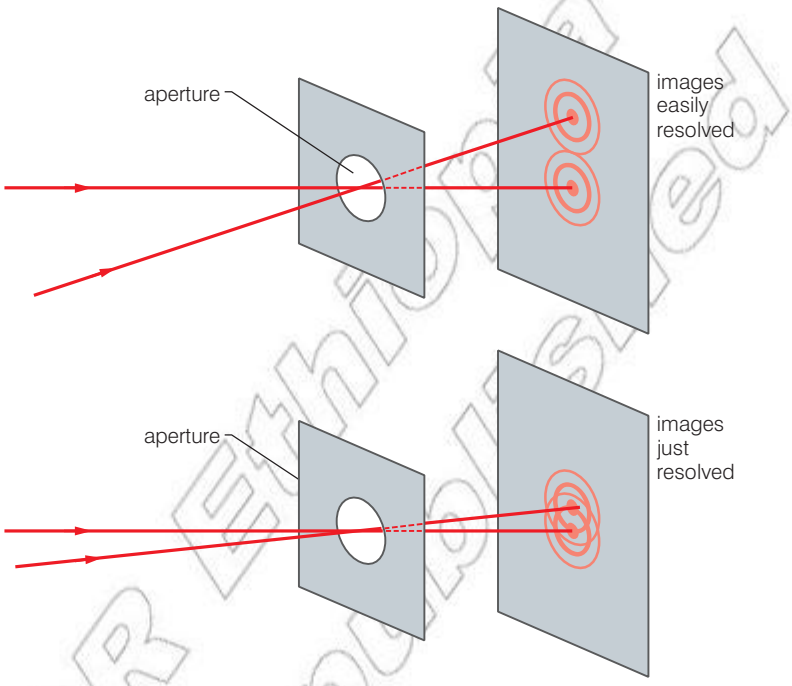


Figure 3.53 *If the interference patterns overlap it becomes very difficult to see the detail or even observe that there may be two different sources of light.*

Think about this...

You may notice the same effect with your eyes. Light from car headlights far away looks like a single point of light. As the car gets closer it becomes more obvious that there are in fact two sources of light.

When using optical microscopes diffraction limits the resolution to approximately 0.2×10^{-6} m. It is not possible to observe any details smaller than this as the interference pattern caused by the diffraction of light blurs the images.

The diffraction grating

As we have seen, light passing through a single narrow slit produces an image which consists of a bright central band with less intense bands on either side of it. Replacing the single slit by two parallel narrow slits produces the same diffraction pattern as the single slit, but in addition it is crossed by a series of interference bands. What happens if further slits are added? Figure 3.54 shows how the pattern changes – note that the slit spacing in each case is the same.

Three parallel slits produce a pattern similar to that produced by two slits, with two important differences. In the three-slit pattern the principal maxima are narrower, and a subsidiary maximum is introduced between each pair of principal maxima. As more and more parallel slits are added, the principal maxima decrease in width. At the same time the number of subsidiary maxima increases and their intensity decreases.

A diffraction grating consists of a set of many evenly spaced slits, in which the slit separation is very small. This means that the principal maxima are very narrow, and there are so many subsidiary

© MoE. FOR Ethiopia Not to be republished

maxima that they are so faint as to be effectively invisible. A beam of monochromatic light passing through a diffraction grating is split into very narrow maxima, as Figure 3.55 shows. The maxima are numbered outwards from the centre, with the undeviated maximum referred to as the zero order maximum.

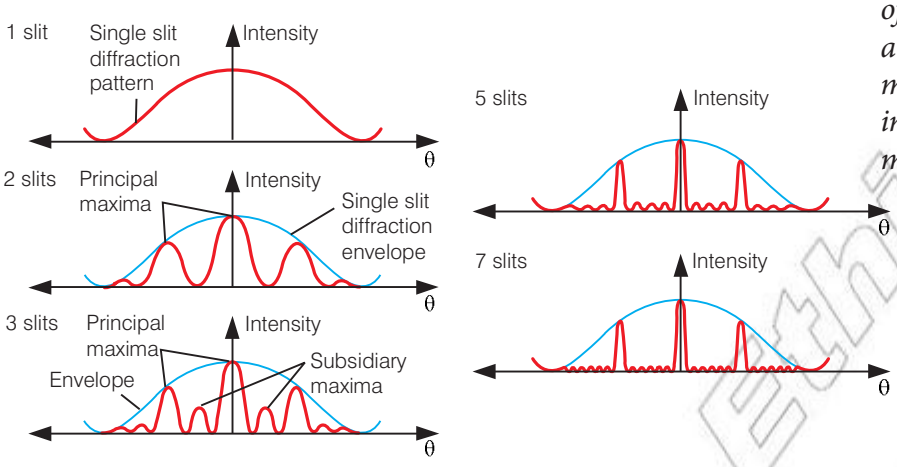


Figure 3.54 *As the number of slits increases the intensity and sharpness of the principal maxima increase while the intensity of the subsidiary maxima decreases.*

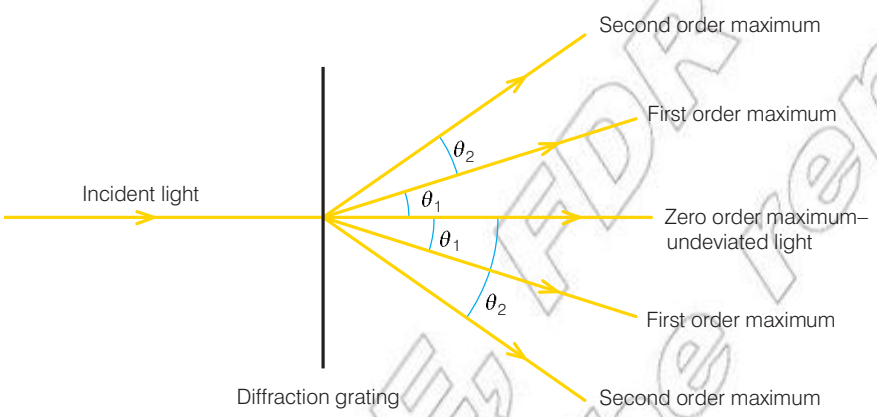


Figure 3.55 *The number of maxima produced by a diffraction grating depends on the wavelength of the light and the distance between the slits of the grating.*

The relationship between the angle at which the maxima occur, the slit separation and the wavelength of light can be obtained as follows. This examination of the diffraction grating assumes that the light strikes the grating with normal incidence.

Each slit in a grating diffracts the incident light, and the diffracted waves then interfere constructively in certain directions only. Figure 3.56 overleaf shows a small portion of a grating. The wavefronts interfere constructively in the direction shown to produce the first order maximum, since the path difference between each adjacent pair of slits is λ. Light from slit A thus interferes constructively with light from slit B (path difference = λ), slit C (path difference = 2λ), slit D (path difference = 3λ), and so on. From Figure 3.56, BF = λ and the slit spacing = AB = *d*.

$$\sin\theta = \frac{BF}{AB} = \frac{\lambda}{d}$$

© MoE/FDR Ethiopia Not to be republished

For the second maximum, BF = 2λ, and:

$$\sin\theta = \frac{BF}{AB} = \frac{\lambda}{d}$$

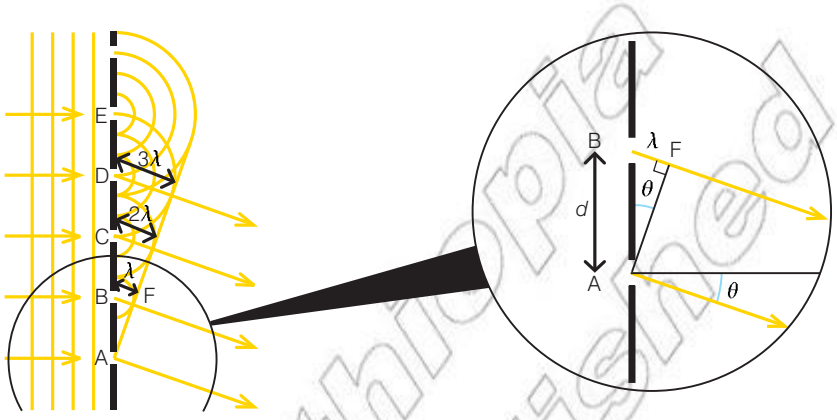


Figure 3.56 *Small portion of a diffraction grating*

In general, the *n*th maximum will occur at an angle θ_n from the zero order maximum, where θ_n is given by:

$$\sin\theta_n = \frac{n\lambda}{d}$$

The spacing of the slits in a grating is sometimes expressed in terms of the number of slits per metre. For a grating with *N* slits per metre, the slit spacing is N^{-1}.

To find out the highest order of the principal maxima, we can use the fact that the maximum value of sin θ is 1. This means that we can write:

$$\frac{n\lambda}{d} \leq 1$$

so:

$$n \leq \frac{d}{\lambda}$$

Since *n* must be a whole number, to calculate the highest order spectrum we calculate the value of d/λ and round it down to the next whole number.

© MoE: FDR Ethiopia Not to be republished

Worked example 3.2

Grating calculations

Yellow light with a wavelength of 5.89×10^{-7} m strikes a diffraction grating with normal incidence. The grating has 5000 slits per centimetre (that is, 5000×100 slits per metre). At what angles will the maxima be seen?

The first order maximum will be when $n = 1$. Since $d = 1/(5000 \times 100)\ \text{m}^{-1}$ we can write:

$$n\lambda = d \sin \theta_n$$

$$1 \times 5.89 \times 10^{-7}\ \text{m} = \frac{1}{(5000 \times 100)\ \text{m}^{-1}} \times \sin \theta_1$$

so:

$$\sin \theta_1 = 1 \times 5.89 \times 10^{-7}\ \text{m} \times (5000 \times 100)\ \text{m}^{-1}$$

$$= 0.2945$$

Therefore $\theta_1 = 17.1°$.

Similarly for θ_2:

$$2 \times 5.89 \times 10^{-7}\ \text{m} = \frac{1}{(5000 \times 100)\ \text{m}^{-1}} \times \sin \theta_2$$

so:

$$\sin \theta_2 = 2 \times 5.89 \times 10^{-7}\ \text{m} \times (5000 \times 100)\ \text{m}^{-1}$$

$$= 0.589$$

Therefore $\theta_2 = 36.1°$.

In the same way, we can show that a third order maximum appears at an angle of 62°. If the calculation for a *fourth* order maximum is carried out, however, we get a value of 1.178 for $\sin \theta_4$. Since the maximum value that the sine of an angle can have is 1, there is no fourth order maximum visible. This can be shown using the value of d/λ:

$$\frac{d}{\lambda} = \frac{(1/(5000 \times 100)\ \text{m}^{-1})}{5.89 \times 10^{-7}\ \text{m}} = 3.40$$

The highest order maximum visible is thus the third order maximum.

Summary

In this section you have learnt that:

- Light passing through a single slit diffracts and creates an interference pattern.
- For interference from a single slit, $a \sin \theta = n\lambda$.
- Diffraction effects limit the details that may be resolved by telescopes and microscopes.
- For a diffraction grating $\sin \theta_n = \frac{n\lambda}{d}$

Ethiopia published

Review questions

1. With the aid of diagrams show how $a \sin \theta = n\lambda$.
2. Compare the interference pattern produced by a single slit for both green and red light.
3. Explain why it is possible to observe greater detail when using blue light in place of red light to illuminate a microscope slide.

End of unit questions

1. Outline Huygens's principle and use it to explain wave propagation, reflection, refraction and diffraction.
2. Describe an experiment to produce a sustained interference pattern of light using a double slit. Explain how the pattern changes if one of the slits is covered over.
3. Describe Young's double slit experiment including an explanation of the purpose of the monochromatic filter and the single slit, and typical values for the slit spacing and distance from slits to screen.
4. In a two slit experiment using light of a wavelength 5.89×10^{-7} m the distance between the two slits and the screen is 1.2 m and the spacing of the slits is 0.55 mm. Calculate the fringe width.
5. A double slit interference experiment is set up in a laboratory using a source of blue monochromatic light of wavelength 475 nm. The separation of the two slits is 0.40 mm and the distance from the slits to the screen where the fringes are observed is 2.20 m.
   a) Describe the interference pattern formed on the screen.
   b) Calculate the fringe separation, and the angle between the middle of the central fringe and the middle of the second bright fringe.
6. Red light from a laser is passed through a single narrow slit and a pattern of bright and dark fringes can be seen on a screen placed 5 m from the slit.
   a) Sketch a graph showing the variation of intensity across the interference pattern.
   b) Describe the effect on the pattern if the width of the narrow slit is reduced.
7. A screen is placed 40.0 cm from a single slit, which is illuminated with light of wavelength 680 nm. If the distance between the first- and third-order minima in the diffraction pattern is 3.0 mm, what is the width of the slit?

© MoE, Brunei Darussalam
Not to be reproduced

Electrostatics

Unit 4

Contents

Section	Learning competencies
4.1 Electric charge and Coulomb's law (page 143)	• Analyse, in quantitative terms, electric fields and the forces produced by a single point charge, two point charges and two oppositely charge parallel plates. • Define the term electric dipole and electric dipole moment. Describe what happens to a dipole placed inside an electric field. • State Gauss' law and define Gaussian surface and electric flux. • Describe and explain in quantitative terms the electric field that exists inside and on the surface of a charged conductor. • Describe Millikan's oil drop experiment.
4.2 Electric potential (page 162)	• Apply the concept of electric potential energy to a variety of contexts. • Use the formula for electric potential due to an isolated point charge. • Apply the concepts of electrical energy to solve problems relating to conservation of energy. • Derive the relationship between electric field strength and potential. • Compare electric potential energy with gravitational potential energy.
4.3 Capacitors and dielectrics (page 173)	• Derive the formula for a parallel plate capacitor (from Gauss' law), including use of a dielectric. • Define the dielectric constant. • Explain qualitatively the charge and discharge of a capacitor in series with a resistor. • Explain the behaviour of an insulator in an electric field. • Define electric energy density and derive the formula for the energy density for an electric field using a parallel plate capacitor. • Solve problems involving capacitances, dielectrics and energy stored in a capacitor.

We have already studied electrostatics; it is the branch of science that deals with the phenomena arising from stationary or static electric charges.

The Ancient Greeks knew that a piece of amber would attract small pieces of straw or hair when rubbed. Scientists such as Benjamin Franklin have also added to our modern understanding of charge. The study of static electricity is called electrostatics.

Using current ideas of physics, charging by friction is understood by thinking of the process of charging with reference to the structure of atoms. Generally speaking, atoms are electrically neutral – that

is to say, they contain equal numbers of positive charges (on the protons in the nucleus) and negative charges (on the electrons around the nucleus). When, say, a glass rod is rubbed by a piece of silk, the glass becomes positively charged while the silk becomes negatively charged. In this process, the silk 'rips off' some of the electrons from the surface of the glass, although how this happens is very poorly understood. It seems that other factors also play a part in the process of charging by friction, since if glass is rubbed with an absolutely clean piece of silk, the glass becomes negatively charged. Even air may have an effect, since experiments show that platinum rubbed with silk in a vacuum becomes negatively charged, whereas it becomes positively charged in air.

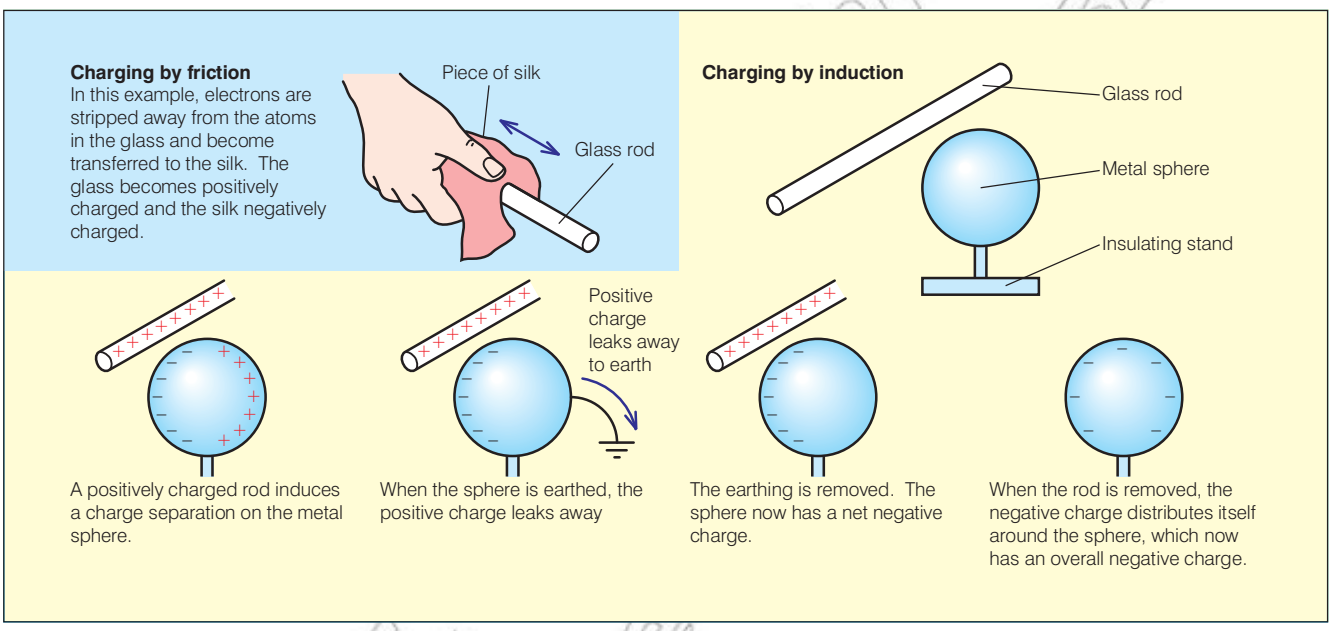


Figure 4.1 *Once an object (in this case a glass rod) has been charged by friction, it may be used to charge other bodies by induction.*

Electrostatic phenomena arise from the forces that electric charges exert on each other. These forces are described by Coulomb's law – this bears several similarities to Newton's laws of gravitation. This unit explores the forces between static charges and the associated energy changes involved when charges are moved from one place to another.

This includes a detailed analysis of capacitors and capacitance. Without these amazing little electronic components most modern electronic equipment would not function, form simple radios to powerful supercomputers.

© MoE Not to be republished

4.1 Electric charge and Coulomb's law

By the end of this section you should be able to:

- Analyse, in quantitative terms, electric fields and the forces produced by a single point charge, two point charges and two oppositely charge parallel plates.
- Define the term electric dipole and electric dipole moment. Describe what happens to a dipole placed inside an electric field.
- State Gauss' law and define Gaussian surface and electric flux.
- Describe and explain in quantitative terms the electric field that exists inside and on the surface of a charged conductor.
- Describe Millikan's oil drop experiment.

What are electric fields?

An **electric field** is the region around a charged object where another charged object will experience a **force**.

There are two kinds of charge, **positive** and **negative**. This is referred to as the **polarity** of a charge. Most objects are electrostatically neutral as they contain an equal number of positive and negative charges.

The force between charges may be attractive or repulsive depending on the nature of the charges.

- Opposite charges attract (+ and – or – and +).
- Like charges repel (+ and + or – and –).

Whenever we represent a region of space containing an electric field we draw a number of **electric field lines** (sometimes called lines of flux or lines of force). These lines tell us a number of things about the field.

From the images in Figure 4.3–4.8 overleaf you can see a number of important features of electric field lines and the corresponding electric field.

Direction

Lines of electric flux have a direction. They move away from positive and towards negative. You can consider the direction as the direction of force acting on a small positive test charge. This is repelled from positive and attracted to negative. When drawing field lines consider the path an initial stationary positive test charge would move.

KEY WORDS

electric field *a region of space around a charged object which exerts a force on other charged objects*
force *the capacity to do work or cause physical change*
positive *the charge on a body which has a deficiency of electrons*
negative *the charge on a body which has a surplus of electrons*
polarity *the condition of being positive or negative*
electric field lines *lines representing an electric field in a region of space*

Think about this...

Most objects develop an electrostatic charge by either gaining or losing electrons. Gaining two electrons would give a net charge of –2e or -3.2×10^{-19} C. How many electrons would an object need to lose in order to have a charge of 1 C?

© Mofet Institute Not to be reproduced



Figure 4.2 *Electric fields in the atmosphere can produce some dramatic effects.*

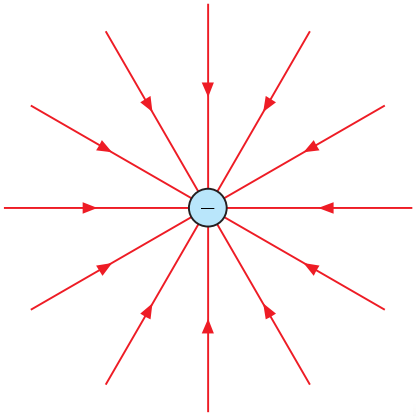


Figure 4.3 *Electric field around a negative point charge*

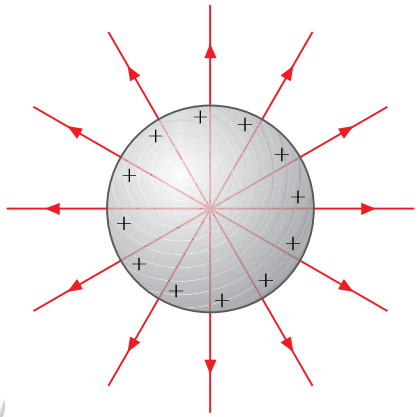


Figure 4.4 *Electric field around a positively charged metal sphere. There is no field inside a metal sphere; the field outside the sphere is the same as it would be if all the charge were concentrated at its centre.*

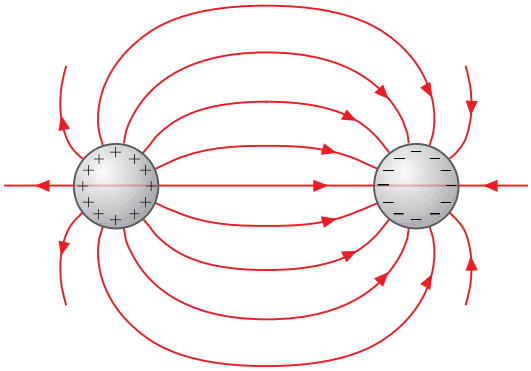


Figure 4.5 *Electric field near one positively charged and one negatively charged small metal sphere*

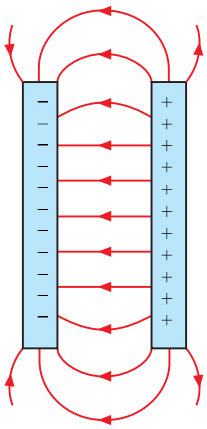


Figure 4.6 *Electric field near two parallel plates, one charged positively and the other negatively. The field in between the plates, but not near their edges, is nearly constant.*

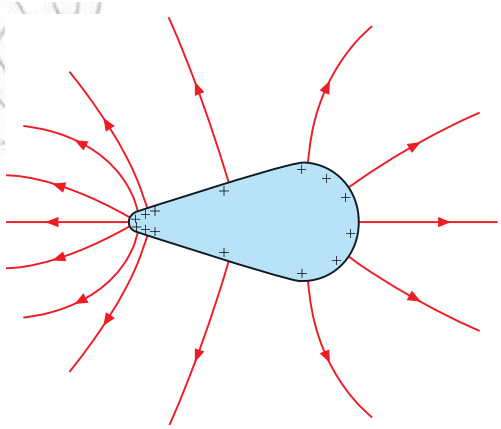


Figure 4.8 *Electric field around a positively charged object rounded at one end and sharp at the other. Electrodes with sharp points are used to provide the strong electric fields in small volumes needed in plasma discharge flat-screen TV displays.*

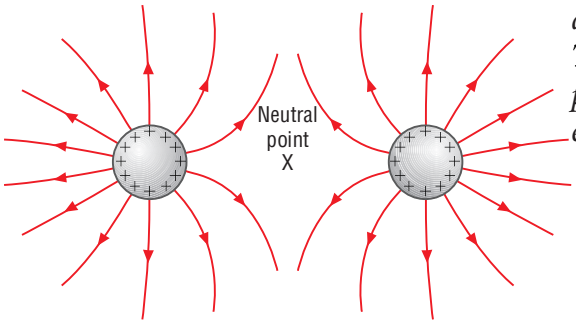


Figure 4.7 *Electric field near two positively charged small metal spheres*

DID YOU KNOW?

The term **test charge** is often used in electrostatics. It is simply a small point charge. In most cases we refer to a positive test charge; the polarity is very important. If it were a negative test charge, the forces would all be reversed.

Crossing

When drawing field lines they must not cross over each other.

Spacing

The spacing of the field lines represents the strength of the field. The closer the lines are together the stronger the field. Look at Figure 4.8; you can see the field is a different strength at different points around the object.

Neutral point

It is possible for two (or more) electric fields to cancel each other out and create a neutral point. Here the resultant field strength is zero.

Activity 4.1: Electric field lines

Figure 4.9 shows a positively charged metal sphere held above an earthed metal plate, that is, held at 0 V.

Copy the diagram and draw at least five electric field lines between the sphere and the plate.

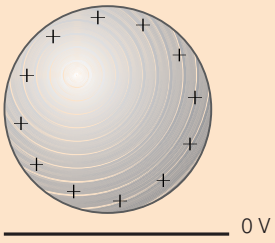


Figure 4.9

KEY WORDS

electric field strength *the force per unit positive charge acting on a positive test charge placed in the field*
vector *a quantity specified by its magnitude and direction*

Electric field strength

At a point within an electric field there is a certain **electric field strength**. The strength of the electric field is defined as:

- The force per unit positive charge acting on a positive test charge placed in the field.

Mathematically this may be expressed as:

- $E = \frac{F}{q}$

E = electric field strength in N/C.

F = force acting on the positive test charge in N.

q = charge of the positive test charge in C.

An electric field strength of 300 N/C literally means a charge of 1 C will experience a force of 300 N.

The equation for electric field strength is often used to determine the force acting on a charged particle due to an electric field:

- $F = Eq$

This has many applications and may be combined with $F = ma$ and the equations of constant acceleration.

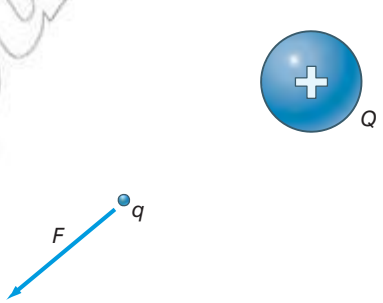


Figure 4.10 *A larger charge (Q) will repel a positive test charge with a force equal to Eq. Notice that q is used to represent the test charge.*

Think about this...

Electric field strength is a **vector** quantity. It is important to include the direction of the lines of force/flux in any field diagrams.

Activity 4.2: Electric field strength

An electron experiences a force of 6.0 μN when passing through an electric field. Calculate the electric field strength.

© Mpumalanga Department of Education
Not to be republished

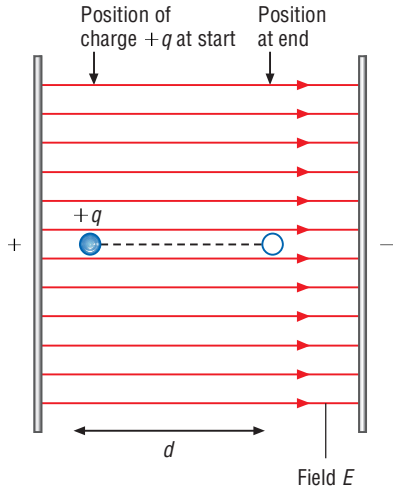


Figure 4.11 *A positive particle moving in a vacuum in the direction of the electric field (or a negative particle moving in the opposite direction to the field)*

Charge movement in the direction of the field

This is the simplest arrangement, as shown in Figure 4.11.

A particle with charge q starts from rest at the point shown. It moves a distance d through the field of electrical field strength E. While it is in the field, a force Eq is acting on it in the direction of the field and therefore work is done on it equal to Eqd.

Since it is in a vacuum its speed will increase and its kinetic energy will rise by Eqd.

Depending on what other information is available (its mass, for example), it would be possible to find its acceleration, its velocity after travelling any particular distance and the time it took to complete that distance. A negatively charged particle travelling in the opposite direction would have the same increase in kinetic energy. A positive particle travelling in the opposite direction would be losing kinetic energy instead of gaining it.

These ideas have considerable practical applications in various types of electronic equipment, for example, in X-ray tubes, particle accelerators, and cathode-ray tubes in oscilloscopes and older television sets, as shown in the following worked example.

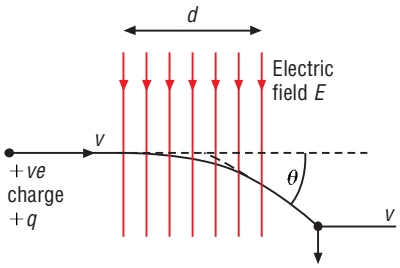


Figure 4.12 *Electrons deflected by an electric field*

Charge movement initially at right angles to the direction of the electric field

Consider a positively charged particle q travelling horizontally with velocity v in a vacuum and entering a uniform electric field of magnitude E for a distance d, as shown in Figure 4.12.

The force that the field exerts on the charge will always be in the direction of the field, so there cannot be any alteration to its horizontal velocity. When the particle emerges from the field, it will have a horizontal velocity equal to its original velocity. It will therefore spend a time d/v in the field. During this time the constant force on it is Eq downwards, giving it a constant acceleration downwards of Eq/m, where m is the mass of the particle.

When it leaves the field, the particle will have a component of velocity downwards of $at = \frac{Eq}{m} \times \frac{d}{v} = \frac{Eqd}{mv}$.

The angle of deflection θ will be given by

$$\tan\theta = \frac{Eqd}{mv^2}$$

© Not to be republished

Activity 4.3: Electron beam

Copy Figure 4.13, which shows an electron beam passing between a pair of parallel conducting plates.

(a) Label the plates positive and negative.

(b) Draw arrows to indicate the relative magnitudes and directions of the force on the electron beam at points X and Y.

(c) The electric field strength between the plates is 8.0×10^4 N/C. Calculate the magnitude of the force on an electron at point X.

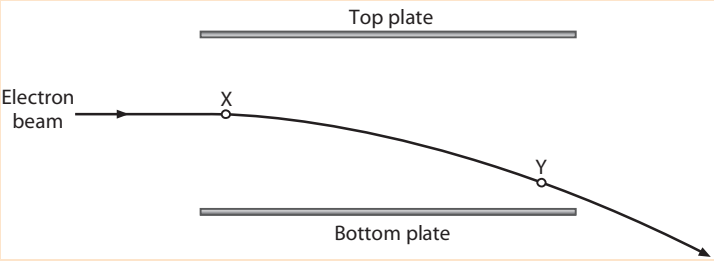


Figure 4.13

Coulomb's law

In the mid-18th century, several scientists were investigating the factors affecting the force between charged particles. The French military engineer Charles Augustin de Coulomb used a torsion balance of his own design to obtain a series of very precise measurements.

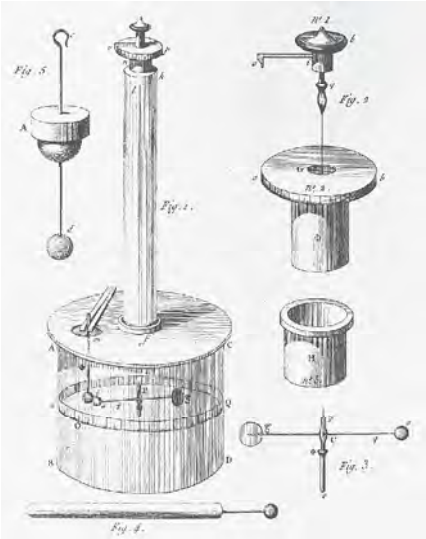


Figure 4.14 *Coulomb's apparatus. A small charged ball a is fixed to a horizontal beam q suspended from a vertical fibre f. On the other end of the beam is an uncharged counterweight g. A second charged ball t is brought up to a, when the electrostatic force rotates the fibre. The force between the balls can be calculated from the angle through which the fibre twists and the torsional constant of the fibre.*



Figure 4.15 *Charles Augustin de Coulomb*

He found two important factors affecting the forces between two charges. In words he found that the force between two charges was:

- proportional to the product of the charges. If the product of the two charges doubled, then the force between them would also double. Mathematically: $F \propto Q_1 Q_2$
- inversely proportional to the square of the distance between the charges. The force between the two charges varies as an inverse

Namibia Not to be republished

square relationship. As you double the distance the force will fall by 4 (2^2). Mathematically:

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

He combined these statements into **Coulomb's law**:

- $F \propto \frac{Q_1 Q_2}{r^2}$

where Q_1 and Q_2 are the two charges and r is the separation between the charges. In most cases the charges are point charges but if we deal with charged spheres the distance must be from the centre of each sphere.



Figure 4.16 *Two positive charges will repel each other with a force proportional to the product of the charges and inversely proportional to the square of the distance between them.*

It follows from Newton's third law that the forces on each charge are equal and opposite. No matter which of the charges is larger they both exert the same repulsive force on each other.

Adding a constant of proportionality to Coulomb's statement gives:

- $F = k \frac{Q_1 Q_2}{r^2}$

The constant k is equal to $1/4\pi\varepsilon_0$ and so this is usually written as:

- $F = \frac{1}{4\pi\varepsilon_0} \frac{Q_1 Q_2}{r^2}$ or $F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$

ε_0 is a constant called the **permittivity of free space** (or vacuum permittivity or electric constant). It has a value of 8.85×10^{-12} F/m. This constant is very important to the study of electric fields. It links electrical concepts such as electric charge to mechanical quantities such as length.

The permittivity of free space may be thought of as a measure of how easy it is for an electric field to pass through a vacuum. Every insulator has a permittivity that is greater than ε_0.

DID YOU KNOW?

Despite the significant progress made by Coulomb he did miss some key details. He explained the laws of attraction and repulsion between electric charges. However, he speculated that the forces were due to different kinds of fluids flowing in the substances.

KEY WORDS

Coulomb's law *law stating that the electrical force between two charged objects is directly proportional to the product of the quantity of charge on the objects and inversely proportional to the square of the separation distance between the two objects.*

permittivity of free space *a constant that specifies how strong the electric force is between electric charges in a vacuum*

DID YOU KNOW?

Along with the permittivity of free space there is a similar constant relating to magnetic fields. This is called the permeability of free space (μ_0). Maxwell equations link these two concepts with the speed of light through a vacuum:

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

This provides clear evidence that light is in fact an electromagnetic wave.

© Ethiopia FDRE Ministry of Education Not to be republished

Worked example 4.1

Determine the force between two protons a distance of 2.0 mm

- $F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$ *State Coulomb's law*
- $F = \frac{1.6 \times 10^{-19} \times 1.6 \times 10^{-19}}{4\pi \times 8.85 \times 10^{-12} \times (2.0 \times 10^{-3})^2}$ *Substitute known values*
- $F = 5.8 \times 10^{-23}$ N. *Solve equation and give units*

This force may seem small but due to the tiny mass of a proton this will give rise to an acceleration of 35 000 m/s!

Activity 4.4: Force calculations

Calculate the force between two point charges of 6.0×10^{-12} C a distance of 9.0 mm apart. Calculate the force between the two charges when: a) one of the charges changes to 9.0×10^{-12} C; b) the distance increases to 12 mm.

The forces between two charges q and Q may be either attractive or repulsive depending on their relative charge. If both charges are the same charge (e.g. positive) they will repel each other. If the two charges are opposite they will attract each other. In each case the magnitude of the force is given by Coulomb's law. However, you will need to consider the direction carefully.

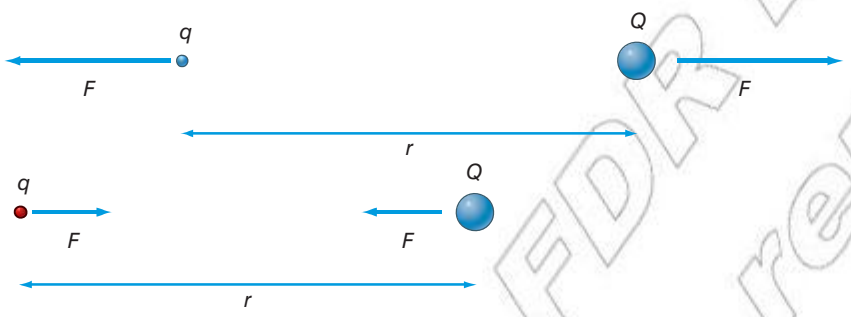


Figure 4.17 *Coulomb's law only gives the magnitude of the forces between charges. Their polarity determines the direction of the force. It is worth producing a simple sketch of the two charges involved in order to easily determine the direction of each force.*

How can we determine the forces due to multiple charges?

If there are multiple charges they will all exert a force on each other. Take the simple example of three charges in a line.

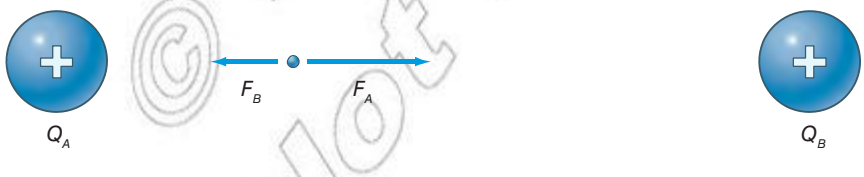


Figure 4.18 *Three charges along a one-dimensional line in space. The test charge will experience a force from both the other charges. The resultant force will be the vector sum of these forces.*

The small test charge q will experience a force due to Q_A and a force due to Q_B. In this example all charges are positive so the forces will be repulsive and as q is between the two charges the forces will be acting opposite directions. Equally the charge Q_A will experience

© NSBT FDR Education Not to be republished

Activity 4.5: Resultant force

The test charge is 12 mm from Q_A and 20 mm from Q_B. If Q_A has a relative charge of $+4e$ and Q_B has a relative charge of $+8e$ determine the magnitude and direction of the result force acting on the test charge.

two repulsive forces, on from charge q and the other from Q_B.

The resultant force acting on charge q will be given by:

- $F_{net} = F_A - F_B = \frac{Q_A q}{4\pi\varepsilon_0 r_A^2} - \frac{Q_B q}{4\pi\varepsilon_0 r_B^2}$

or

- $F_{net} = \frac{q}{4\pi\varepsilon_0}\left(\frac{Q_A}{r_A^2} - \frac{Q_B}{r_B^2}\right)$

If Q_B were to be replaced by a negative charge then both the forces acting on the test charge would be acting in the same direction so the resultant force will be the sum of the two forces.

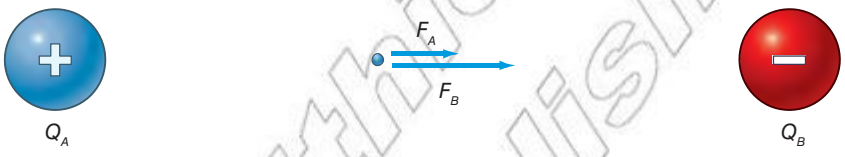


Figure 4.19 *If the forces due to both charges act in the same direction the resultant force will be the sum of* F_A *and* F_B.

A similar process will enable the resultant force to be determined if the charges are not limited to one dimension.

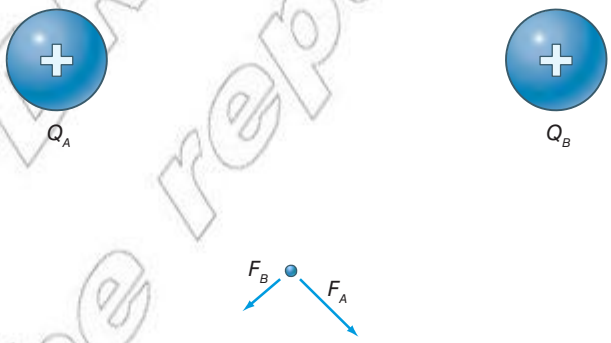


Figure 4.20 *The test charge will experience two repulsive forces from the positive charges. These force acts outwards, away from the positive charge.*

In this case the resultant force will need to be determined using vector addition.

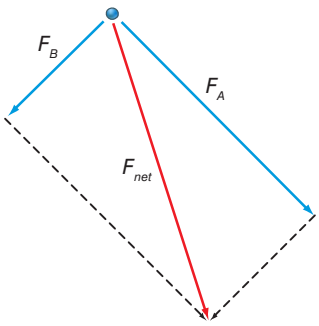


Figure 4.21 *The resultant force may be determined using vector addition (either scale diagrams of mathematical addition using resolving, trigonometry and Pythagoras's' theorem).*

© MoE, FDR Ethiopia Not to be republished

Experimental verification of Coulomb's law

Coulomb's law may be verified experimentally using reasonably simple apparatus. Measurements must be taken quickly, in order to avoid the problem of charge leaking away. Figure 4.22 shows the details of the investigation.

From the free body diagram it can be seen that when the ball is in equilibrium:

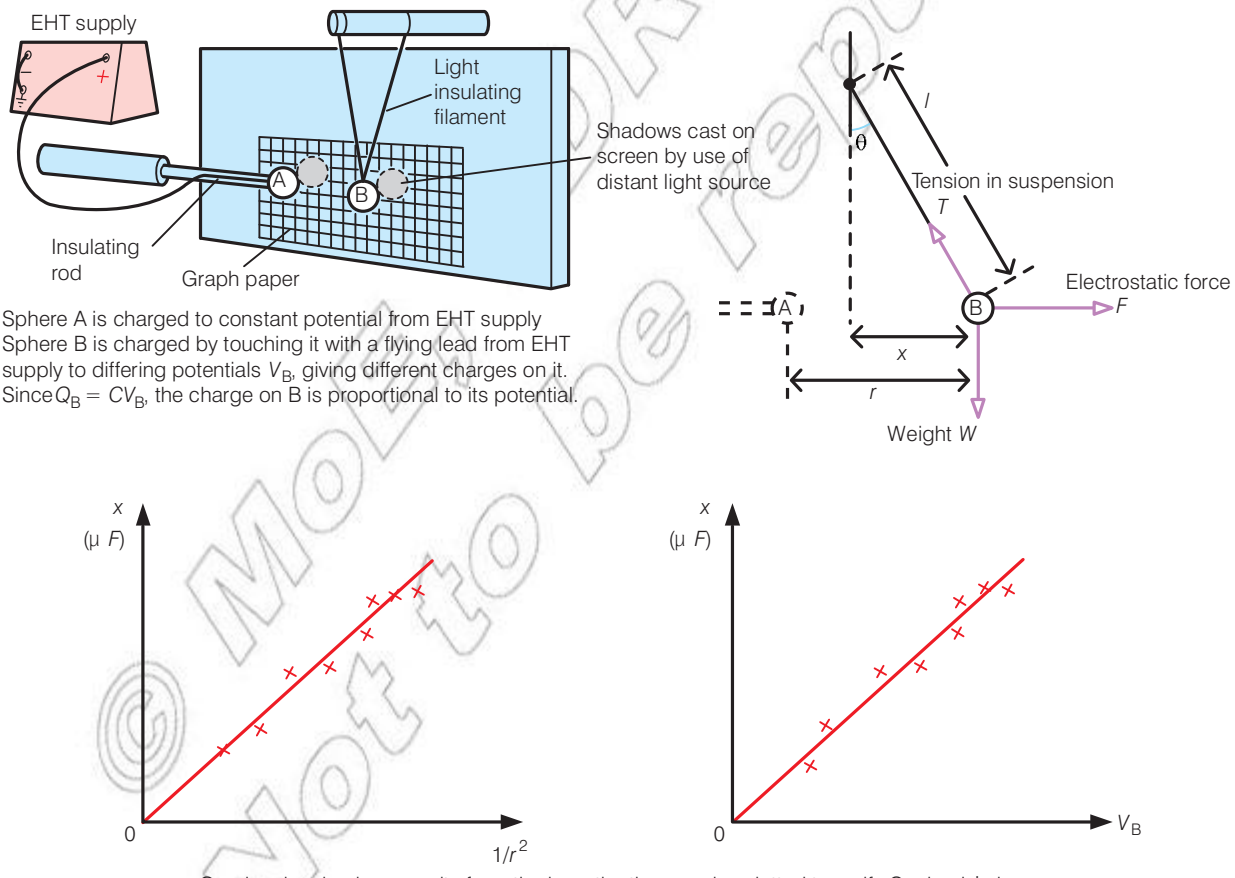
$$F = T \sin\theta \quad \text{and:} \quad W = T \cos\theta$$

Dividing the first equation by the second we get:

$$\frac{F}{W} = \frac{T \sin\theta}{T \cos\theta} = \tan\theta \quad \text{so:} \quad F = W \tan\theta$$

Now $\tan\theta \approx \frac{x}{l}$ if θ is small, so: $F = \frac{Wx}{l}$

Thus for small deflections, the horizontal deflection of the ball from its equilibrium position (from its position when no other charged body is present) is directly proportional to the electrostatic force F. Confirmation of Coulomb's law is obtained from graphs like those shown in the figure.



Sphere A is charged to constant potential from EHT supply
Sphere B is charged by touching it with a flying lead from EHT supply to differing potentials V_B, giving different charges on it.
Since $Q_B = CV_B$, the charge on B is proportional to its potential.

Graphs showing how results from the investigation may be plotted to verify Coulomb's law

Figure 4.22 *Apparatus to verify Coulomb's law. Very fine nylon thread may be used to suspend the ball. The balls are usually made of expanded polystyrene painted with aluminium paint to give them a conducting layer. Glass is suitable for the insulating rod.*

© MoE-EMDR Ethiopia Not to be republished

How do electrostatic forces compare to gravitational forces?

There are several similarities between electric and gravitational fields. Both use similar mathematics to describe their effects. However, there are also a few key differences.

- Electrostatic forces may be attractive or repulsive, gravitational forces are only ever attractive.

Electrostatic forces are significantly stronger than gravitational forces (see below).

Worked example 4.2

Calculations suggest that there are around 10^{29} electrons and the same number of protons in the body of a 70 kg person. Taking the charge on an electron as -1.6×10^{-19} C, calculate the magnitude of the electrostatic force F_e between the electrons in one 70 kg person and the protons in another 70 kg person standing 2 m away. Compare this with the size of the gravitational force F_g between the two people.

Total charge on electrons in first person = $10^{29} \times -1.6 \times 10^{-19}$ C = -1.6×10^{10} C

Total charge on protons in second person = $10^{29} \times +1.6 \times 10^{-19}$ C = $+1.6 \times 10^{10}$ C

So the magnitude of the electrostatic attractive force F_e on each person is:

$$F_e = \frac{1}{4\pi\varepsilon_0}\frac{Q_1 Q_2}{r^2}$$

$$= \frac{1}{4\pi \times 8.854 \times 10^{12}\ \text{F/m}} \quad \frac{(1.6 \times 10^{10}\ \text{C}) \times (1.6 \times 10^{10}\ \text{C})}{(2\ \text{m})^2}$$

$$= 5.75 \times 10^{29}\ \text{N}$$

By comparison, the magnitude of the gravitational attractive force F_g on each person is:

$$F_g = \frac{Gm_1m_2}{r^2}$$

$$= \frac{6.67 \times 10^{11}\ \text{N m}^2/\text{kg}^2 \times 70\ \text{kg} \times 70\ \text{kg}}{(2\ \text{m})^2}$$

$$= 8.2 \times 10^{-8}\ \text{N}$$

The electrostatic force is nearly 10^{37} times bigger than the gravitational force, although of course the attractive electrostatic force between the electrons and protons is cancelled exactly by the repulsive force between the two sets of protons (or the two sets of electrons).

KEY WORDS

Gauss's law *law stating that the electric flux though any closed surface is proportional to he enclosed electric charge*
electric flux *a measure of the number of electric field lines passing through a surface*

Gauss's law and electric field strength

Gauss's law (also known as Gauss's flux theorem or simply Gauss's theorem) describes the link between the distribution of charge on an object and its resulting electric field.

The electric flux through a surface is defined as the electric field multiplied by the area of the surface perpendicular to the field:

- φ = area × electric field strength
- $\varphi = AE$

Gauss's law may be written as:

- The **electric flux** through any closed surface is proportional to the enclosed electric charge.

It can be shown that the total electric flux out of a closed surface is equal to the charge enclosed divided by the permittivity:

- $\varphi = \frac{Q}{\varepsilon_0}$

Putting these two equations equal to each other and we get:

- $AE = \frac{Q}{\varepsilon_0}$

Therefore the electric field strength around a charged object is given by:

- $E = \frac{Q}{A\varepsilon_0}$

It is often useful to construct an imaginary surface (called a Gaussian surface). This enables simple calculations to determine the field strength at any given point on the surface. As long the shape of the surface is simple (e.g. sphere, cylinder, etc.) then Gauss's law greatly simplifies calculation of the electric field strength. Gauss's law can also be used to derive Coulomb's law and vice versa.

There are three other consequences of Gauss's law and Coulomb's law concerning the distribution of charges on a charged conductor:

1. **The net electric charge of a conductor resides entirely on its surface**. This is due to the repulsion of like charges; the charges are pushed as far apart as possible and so spread out on the surface.
2. **The electric field inside the conductor is zero**. If there were to be any charge inside the object then this would cause there to be a net force acting on some of the charges and they would accelerate. This is not the case, the charges remain static.
3. **The electric field at the surface of the conductor is perpendicular to that surface**. If the field were to act at an angle then there would be a horizontal component of the force. This would again cause the charges to move around rather than remaining static.



Figure 4.23 *Gauss was an outstanding mathematician.*

DID YOU KNOW?

Johann Carl Friedrich Gauss was a German mathematician born in 1777. He is often referred to as the Princeps mathematicorum – this is Latin for the Prince of Mathematicians. He had a remarkable influence in many fields of mathematics and science including optics, astrophysics and number theory.

Electric field around a point charge

If we consider a point charge (Q) the electric field strength at a given distance (r) may be found by:

- $E = \frac{Q}{4\pi\varepsilon_0 r^2}$ or $E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}$

This equation may be derived is two ways.

1. From Coulomb's law and electric field strength, i.e. from $E = F/q$ and substituting.

Activity 4.6: Electric field strength of a proton

Taking values between 0.1 and 1.0 mm plot a graph of electric field strength against distance from a proton. Verify that this is an inverse square relationship.

© MDE Ethiopia Not to be republished

- $F = \frac{Qq}{4\pi\varepsilon_0 r^2}$.

2. From Gauss's law. The surface area of the sphere is equal to $4\pi r^2$. Therefore at a distance r the area through which the total flux must pass through is equal to $4\pi r^2$ – this is an example of a simple Gaussian surface.
   $E = \frac{Q}{A\varepsilon_0}$ becomes $E = \frac{Q}{4\pi r^2\varepsilon_0}$.

The electric field strength around a point charge varies as an inverse square relationship. As you double the distance the field strength will fall by 4 (2^2). Mathematically: $E \propto \frac{1}{r^2}$ and so $E = \frac{k_2}{r^2}$ or $Er^2 = k_2$.

Electric field between two parallel plates

Think about this...

It is tempting to assume that the electric field is stronger near the plates. However, this is not true. The field has the same strength everywhere in between the two plates.

An electric field is formed in the region between two oppositely charged parallel plates. Unlike the examples we have looked at so far, the field in this region is uniform.

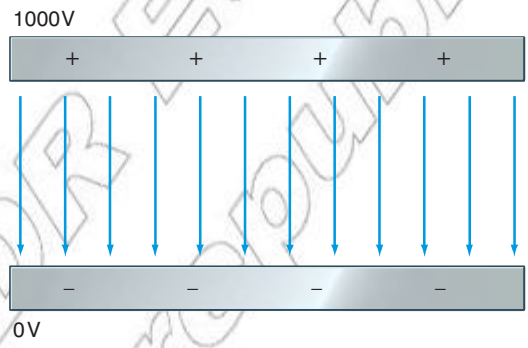


Figure 4.24 *The field between two oppositely charged parallel plates is uniform. This means the field lines are equal spaced between the plates.*

Using Gauss's law we can derive an expression for the electric field strength between the plates.

KEY WORDS

charge density *the charge per unit area on a charged surface*

The charge density on each plate is the charge per unit area (in C/m^2). For a given plate with a charge Q and an area A the **charge density** σ is given by:

- $\sigma = Q/A$

Using Gauss's law the electric field through each plate is:

- $E = \sigma/2\varepsilon_o$

As the plates have an equal but opposite charge they mutually attract and hold each other together. Therefore in between the plates the electric fields are in the same direction. The total electric field between the plates is given by the sum of the two electric fields.

- $E = \sigma/2\varepsilon_o + \sigma/2\varepsilon_o$
- $E = 2(\sigma/2\varepsilon_o)$
- $E = \sigma/\varepsilon_o$

This equation is often referred to as the parallel plate capacitor equation (more on this in section 4.3).

© MoE-FDRE Ethiopia Not to be republished

It is also important to note outside of the plates the fields due to each plate are acting in different directions. They cancel out leaving a uniform field inside the plates but not field outside of the plates.

It can also be shown the electric field strength between the two plates can be given by

The electric field strength between the two plates is given by

- $E = \frac{V}{d}$

V = potential difference across the plates in V.
d = distance between the plates in m.

This equation will be explored in more detail in section 4.2. However, it should be noted that this also demonstrates that electric field strength may be measured in V/m in addition to N/C.

Activity 4.7: Fields between plates

Copy the diagram in Figure 4.24. The field between two oppositely charged parallel plates is uniform. This means the field lines are equally spaced between the plates. Draw two other diagrams with:

i) 500 V instead of 1000 V

ii) the plates half the distance apart and the polarity reversed.

In both cases think carefully about the field lines.

By combining the above equation and our defining equation for electric field strength we get:

$E = \frac{V}{d} = \frac{F}{q}$

This is commonly used to determine the force acting on a charged particle between two parallel plates in the form of:

$E = \frac{Vq}{d}$

Activity 4.8: Electron between plates

Calculate the force acting on an electron as it passes between two plates with a p.d. of 500 V and a separation of 40 mm.

Charged particle movement in an electric field

Charged particles can be controlled by a potential difference along their path. If the reverse potential difference is high enough, the particles are stopped and sent backwards. This effect is similar to the gravitational effect when a ball is thrown vertically upwards; when the ball loses all its kinetic energy, it falls back to the ground. We shall deal below with various movements of charge for a variety of directions of electric fields.

© Ethiopia Not to be republished

Whenever a charge particle moves through an electric field it accelerates in the direction of the force acting on it. Therefore if the force is parallel to the direction of motion of the particle it accelerates. Using Newton's second law the acceleration is given by:

$F = ma$ (Newton's second law)

$Eq = ma$ (Substituting F for the force on a charged particle in an electric field)

$a = \frac{Eq}{m}$ (Rearrangeing to make a the subject)

If the force acts perpendicularly then the particle will follow a parabolic path, just like a ball through the air

Worked example 4.3

In a cathode-ray tube, electrons leave a cathode (which is negative) and are accelerated for a distance of 4.0 cm by a uniform electric field of electric field strength 1.20×10^5 N/C. They then pass through a hole in the anode (which is positive) and enter a region in which the electric field strength is zero. Calculate:

a) the speed of an electron when it reaches the anode

b) the time it takes to reach the screen of the cathode-ray tube, that is 28 cm from the anode.

Answer:

a) Electric field strength = V/d, therefore
$V = 1.20 \times 10^5 \times 0.040 = 4800$ V or 4800 J/C.

The charge on an electron is 1.6×10^{-19} C.
Energy gained = 4800 J/C × 1.60×10^{-19} C = 7.68×10^{-16} J.

This is the kinetic energy of the electron, that is,
$7.68 \times 10^{-16} = \frac{1}{2} mv^2 = 0.5 \times 9.11 \times 10^{-31} \times v^2$, giving

$$v = \sqrt{\frac{2 \times 7.68 \times 10^{-16}}{9.11 \times 10^{-31}}} = 4.11 \times 107 \text{ m/s}$$

b) The electron then coasts at constant speed until it reaches the screen.

Time taken = 0.28 m/4.11×10^7 m/s = 6.8×10^{-9} s.

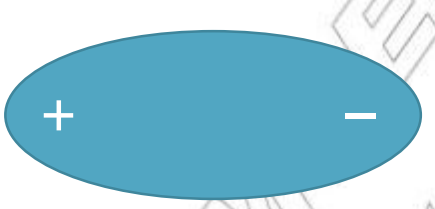


Figure 4.25 *An example of a simple dipole*

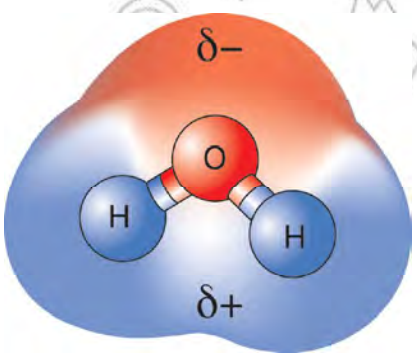


Figure 4.26 *Water forms a permanent dipole.*

Electric fields and dipoles

An electric dipole is created whenever there is a separation of positive and negative charges. This often happens at the molecular level and leads to molecular dipoles. A single atom may have its cloud of electrons moved slightly in one direction by an external electric field. This leads a separation of the charge (the positive nucleus and the negative electrons). This is called an induced electric dipole. The atom only forms a dipole while the electric field is present, switch it off and the atom returns to normal.

Some molecules form permanent dipoles. Perhaps the best example is water. The oxygen side of the molecule is always slightly negative (δ-) and the hydrogen side of the molecule is always slightly positive (δ+).

Activity 4.9: Attracting water

Rub a polythene rod with a duster then place the rod near a thin trickle of water from a tap. The negative static charge built up on the rod will attract the positive side of the water molecule and the trickle will bend towards the rod.

An electric dipole moment for a pair of opposite charges of magnitude q is a measure of the system's overall polarity. It is defined as the magnitude of the charge multiplied by the separation between the two charges.

- *electric dipole moment* = qd
- $P = qd$

It is a vector quantity and its direction is always towards the positive charge. The electric dipole moment helps describe the orientation of the dipole.

When a dipole is placed inside an electric field equal and opposite forces act on the charges in the dipole. This creates a turning effect and so the dipole orientates itself with the electric field.

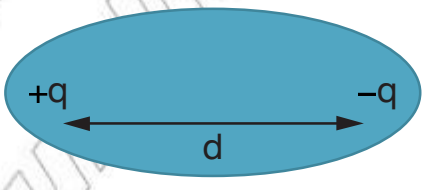


Figure 4.27 *Calculating the dipole moment*

Millikan's oil drop experiment

In experiments to determine the nature of fundamental particles, scientists can use a mass spectrometer to find the ratio of charge to mass for particles. This ratio is a fundamental property of a charged particle and identifies it uniquely, but until the determination of the value of the charge on an electron, it was impossible to break down the ratio and find the mass of that particle. In 1909, Robert Millikan developed an experiment which determined the charge on a single electron.

Although he had a variety of extra bits and pieces to make it function successfully, the basic essence of Millikan's experiment is pure simplicity. The weight of a charged droplet of oil is balanced by the force from a uniform electric field, so that the oil drop remains stationary. Using the same apparatus, Millikan undertook variations, one in which there was no field and the downward terminal velocity of the oil drop was measured, and another in which the field was adjusted to provide a stronger force than gravity and the terminal velocity upwards was measured.

When oil is squirted into the upper chamber from the vaporiser, friction gives the droplets an electrostatic charge. This will be some (unknown) multiple of the charge on an electron, because electrons have been added or removed due to the friction. As the drops fall under gravity, some will go through the anode and enter the uniform field created between the charged plates. If the field is switched off, they will continue to fall at their terminal velocity.

DID YOU KNOW?

Millikan was a professor at the University of Chicago. He was working with a student (Harvey Fletcher) on the oil drop experiment. Despite working together Millikan took sole credit for the discovery of the charge on the electron (which won him the Nobel Prize in 1923).

© FDRE MoE for Ethiopia
Not to be published

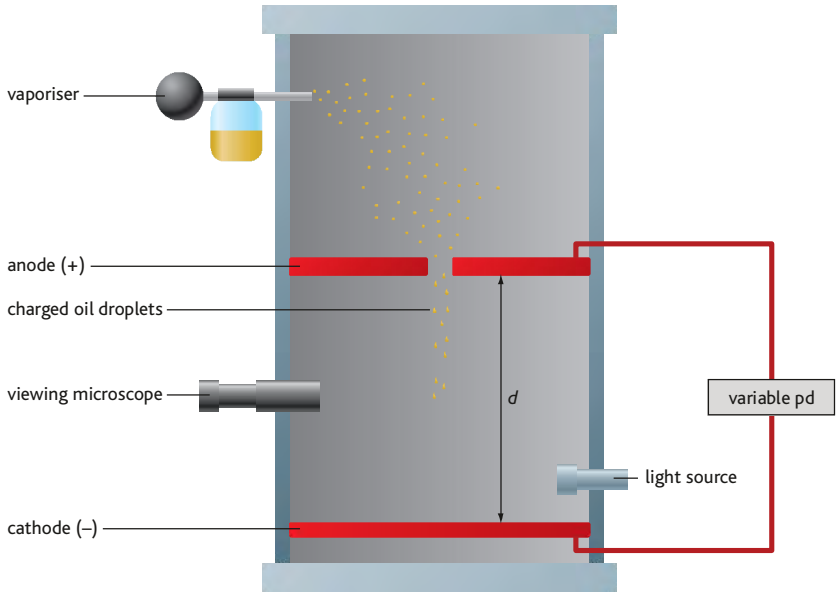


Figure 4.28 *Schematic of Millikan's oil drop chamber*

Stokes' law

Sir George Gabriel Stokes investigated fluid dynamics and derived an equation for the viscous drag (F) on a small sphere moving through a fluid at low speeds:

$$F = 6\pi\eta vr$$

where r is the radius of the sphere, v is the velocity of the sphere, and η is the coefficient of viscosity of the fluid.

For Millikan's oil drops, the density of air in the chamber is so low that the upthrust is generally insignificant (although it would have to be considered if we wanted to do really accurate calculations). At the terminal velocity, the weight equals the viscous drag force:

$$mg = 6\pi\eta v_{term} r$$

where η is the viscosity of air and r is the radius of the drop.

When held stationary by switching on the electric field and adjusting the potential, V, until the drop stands still:

weight = electric force

$$mg = QE$$
$$= \frac{QV}{d}$$

By equating the expressions for weight from the two situations, it is found that:

$$6\pi\eta v_{term} r = \frac{QV}{d}$$

or

$$Q = \frac{6\pi\eta v_{term} rd}{V}$$

© MoE. Free Edition. Not to be reproduced

Millikan could not measure r directly, so had to eliminate it from the equations. Further development of Stokes' law tells us that a small drop falling at a low terminal velocity will follow the equation:

$$v_{term} = \frac{2r^2 g(\rho_{oil} - \rho_{air})}{9\eta}$$

which, if we again ignore the density of air, rearranges to:

$$r = \left(\frac{9\eta v_{term}}{2g\rho_{oil}}\right)^{½}$$

Overall then:

$$Q = \frac{6\pi\eta v_{term} d}{V} \times \left(\frac{9\eta v_{term}}{2g\rho_{oil}}\right)^{½}$$

Millikan did the experiment several hundred times, including repeated measurements on each drop, over and over again letting it fall, halting it with a field, and then lifting it up again with a stronger field, before letting it fall again. From these data, he found that the charges on the droplets were always a multiple of 1.59×10^{-19} C, which is less than 1% away from the currently accepted value of 1.602×10^{-19} C. For this (and work on the photoelectric effect) Millikan was awarded the 1923 Nobel Prize for Physics.

Activity 4.10: Oil drop

Figure 4.29 shows an oil droplet at rest between two conducting plates.

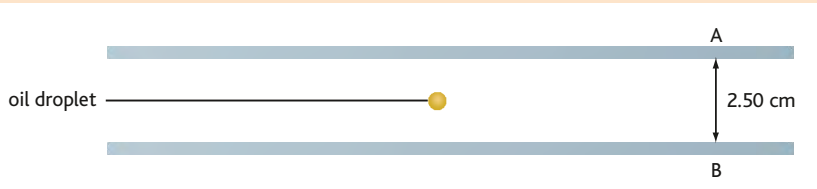


Figure 4.29 *A positively charged oil drop held at rest between two parallel conducting plates A and B*

a) The oil drop has a mass 9.79×10^{-15} kg. The potential difference between the plates is 5000 V and plate B is at a potential of 0 V. Is plate A positive or negative?

b) Draw a labelled free-body force diagram that shows the forces acting on the oil drop.

c) Calculate the electric field strength between the plates.

d) Calculate the magnitude of the charge Q on the oil drop.

e) How many electrons would have to be removed from a neutral oil drop for it to acquire this charge?

DBE Ethiopia Not to be republished

Summary

In this section you have learnt that:

- An electric field is a region of space around a charged object.
- Electric field strength is a vector quantity defined as the force per unit positive charge acting on a positive test charge placed in the field, given by $E = \frac{F}{q}$.
- Coulomb's laws states: The force between two charges is proportional to the product of the charges and inversely proportional to the square of the distance between the charges.

  Expressed mathematically as: $F = \frac{1}{4\pi\varepsilon_0}\frac{Q_1Q_2}{r^2}$ or $F = \frac{Q_1Q_2}{4\pi\varepsilon_0 r^2}$.
- The electric field strength around a point charge is given by $E = \frac{Q}{4\pi\varepsilon_0 r^2}$ or $E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}$.
- Gauss' law states the electric flux through any closed surface is proportional to the enclosed electric charge.
- A Gaussian surface is an imaginary surface around a charge. This enables simple calculations to determine the field strength at any given point on the surface.
- The electric flux through a surface is defined as the electric field multiplied by the area of the surface perpendicular to the field.
- The electric field between two parallel plates is uniform and the field strength may be calculated using using

  $E = \sigma/\varepsilon_0$ or $E = \frac{V}{d}$.
- An electric dipole is created whenever there is a separation of positive and negative charges.
- Dipole moment (P) is given by $P = qd$.
- When a dipole is placed within an electric field the moment leads to a turning effect, orientating the dipole with the electric field.

© MoE Not

Review questions

1. Explain the meaning of the terms electric field and electric field strength.
2. A small conducting sphere is attached to the end of an insulating rod (not shown). It carries a charge of $+5.0 \times 10^{-9}$ C. The sphere is held between two parallel metal plates. The plates, which are 4.0 cm apart, are connected to a 50 000 V supply.

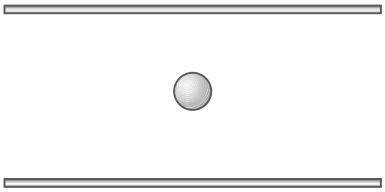


Figure 4.30

Calculate:

a) the magnitude of the electric field strength between the plates

b) the magnitude of the force on the sphere, treated as a point charge of $+5.0 \times 10^{-9}$ C.

3. A second identically charged sphere like that in Question 2 above is attached to a top pan balance by a vertical insulating rod. The charged sphere of Question 2 is clamped vertically above the second sphere such that their centres are 4.0 cm apart.

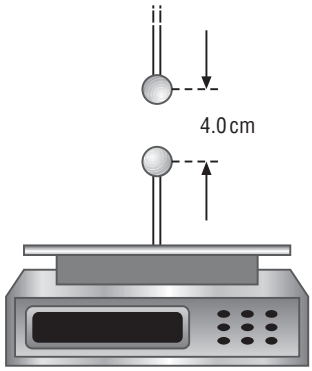


Figure 4.31

a) Show that the force between the two spheres acting as point charges is about 1.4×10^{-4} N.

b) The balance can record masses to the nearest 0.001 g. The initial reading on the balance before the original charged sphere is clamped above the second sphere is 8.205 g. Calculate the final reading in g on the balance.

4. a) Some details of the accelerating plates of an oscilloscope are shown overleaf in Figure 4.32. Electrons leave the cathode with negligible kinetic energy. They are accelerated through a vacuum towards the anode at 0 V.

© FDRE Ethiopia Not to be republished

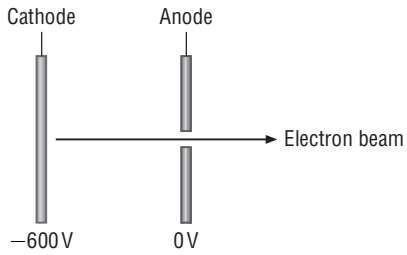


Figure 4.32

Calculate

(i) the kinetic energy and

(ii) the speed gained by the electrons as they pass through the anode.

b) The electron beam passes through a pair of deflecting plates before hitting a fluorescent screen, as shown in Figure 4.33.

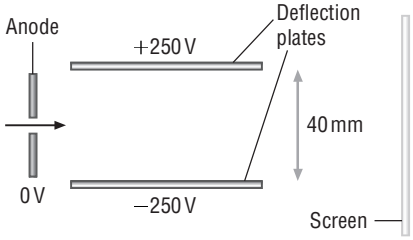


Figure 4.33

i) Calculate the electric field strength between the deflecting plates.

ii) Copy Figure 4.33 and sketch on it:

1 five lines to represent the electric field in the space between the plates

2 the path of the beam from the anode to the screen.

iii) Describe and explain the shape of the path of the beam through the region of electric field.

5. Define Gauss' law and describe the use of a Gaussian surface.

KEY WORDS

work *the amount of energy transferred when an object is moved through a distance by a force*
electrical potential energy *the potential energy acquired by a charged object as it moves in an electric field*
electrical potential *the work done per unit positive charge to move a positive test charge from infinity to its current position within an electric field*

4.2 Electric potential

By the end of this section you should be able to:

- Apply the concept of electric potential energy to a variety of contexts.
- Use the formula for electric potential due to an isolated point charge.
- Derive the relationship between electric field strength and potential.
- Apply the concepts of electrical energy to solve problems relating to conservation of energy.
- Compare electric potential energy with gravitational potential energy.

© MoE; FDRE Ethiopia Not to be republished

What is electrical potential?

In order to bring two positive charges closer to each other **work** has to be done as the charges repel each other. The charges gain **electrical potential energy** rather like an object gains gravitational potential energy when it is lifted vertically in a gravitational field.

If you push one sphere closer to the other one it will gain in electrical potential energy. If you let go, this will be converted into kinetic energy as the mobile sphere rushes away.



Figure 4.34 *Moving two like charges closer to each other requires a force to be applied. Therefore work is done as a force is moved through a distance.*

The change in electrical potential energy is due to a change in **electrical potential** as the sphere moves through the electric field of the larger sphere.

Electric potential has a very specific definition:

- Work done per unit positive charge to move a positive test charge from infinity to its current position within an electric field.

This definition needs some explaining.



Figure 4.35 *The electric potential at A is equal to the work done per unit charge in moving a positive test charge from infinity (i.e. a long way away) to point A.*

The electric field around the large positive charge exerts a repulsive force on the test charge; as a result work is done on the test charge as it moves a distance against a force.

The closer point A is to the large positive charge the greater the amount of work that needs to be done and so the higher the potential.

How do you calculate electric potential?

Electrical potential is a **scalar** quantity and is measured in volts (V). It is usually given the symbol *V*. Care must be taken not to confuse this with electric field strength or electrical potential energy. From the definition electric potential may be calculated using the equation below:

- $V = \frac{W}{q}$

V = electric potential in V or J/C.

W = work done (to move the test charge from infinity to its current position) in J.

q = charge of test charge in C.

You will recall that 1 **volt** is equal to 1 **joule per coulomb**. As a result, an electric potential of 1000 V means that there has been 1000 J of work done per coulomb of charge to move a test charge to its current position within the electric field.

Think about this...

Electrical potential at any point in an electric field may also be defined as the potential energy of each coulomb of positive charge placed at that point.

KEY WORDS

scalar *a quantity specified only by its magnitude*
volt *a measurement of voltage or electromotive force, defined as joules per coulomb*
joule per coulomb *a measurement of voltage or electromotive force, also known as a volt*

DID YOU KNOW?

When using a voltmeter in an electric circuit you are really measuring potential difference – that is to say the difference in electrical potential between two points in the circuit.

© NEW ERA Education Not to be reproduced

If the test charge had a charge of 2 C the work done would be 2000 J, if it had a charge of 0.1 C the work done would be 100 J, etc.

From its definition, the electric potential at infinity is 0 V.

The potential around a single point charge

In order to determine the work done to move a test positive charge to a point in an electric field produced by point charge we can't simply use $W = Fx$ as the force is constantly changing (as per Coulomb's law). As a result, more complex techniques are required. It can be shown that the electric potential at any distance r from a single charge Q is given by:

- $V = \frac{Q}{4\pi\varepsilon_0 r}$ or $V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r}$

The electrical potential at any point around a point charge is directly proportional to the magnitude of the charge and inversely proportional to the distance from the charge.



Figure 4.36 *The electric potential around a point charge depends on the charge Q and the distance from the charge r.*

Think about this...

If charge Q was a negative charge the electric potential would also be negative, e.g. –200 V. Can you explain the importance of this in term of our original definition of potential (consider the work done on the positive test charge).

Worked example 4.4

The electric potential 10 cm from a 1.3×10^{-6} C charged sphere is given by:

- $V = \frac{Q}{4\pi\varepsilon_0 r}$ *State the equation for electric potential*
- $V = \frac{1.3 \times 10^{-6}}{(4\pi \times 8.85 \times 10^{-12} \times 0.10)}$ *Substitute known values*
- V = 117 kV (3 s.t.) kV. *Solve equation and give units*

The electric potential around a point charge varies as an inverse proportional relationship. As you double the distance the potential will halve. Mathematically:

$V \propto \frac{1}{r}$ and so $V = \frac{k_1}{r}$ or $vr = k_1$

© Not to be republished

Activity 4.11: Missing quantities

Calculate the missing quantities using the data below:

Electric potential (V)	Charge (C)	Distance (m)
	3.2×10^{-12}	0.05
	-1.6×10^{-16}	1.0×10^{-4}
300 000	$\times 10^{-18}$	
−2000		0.1

Equipotentials

KEY WORDS

equipotential *a line joining points within a field that have the same potential*

An **equipotential** is a line joining points within an electric field (or indeed any field) with the same potential. These are commonly called equipotential surfaces as they are not just one-dimensional.

If a charge moves along a line of equipotential then its potential remains constant and there is no change in its electrical potential energy. This means all lines of equipotential are perpendicular to the electric field lines.

Uniform field

In a uniform field the lines of equipotential are equidistant parallel lines. The diagram in Figure 4.37 shows the lines of equipotential between two oppositely charged parallel plates.

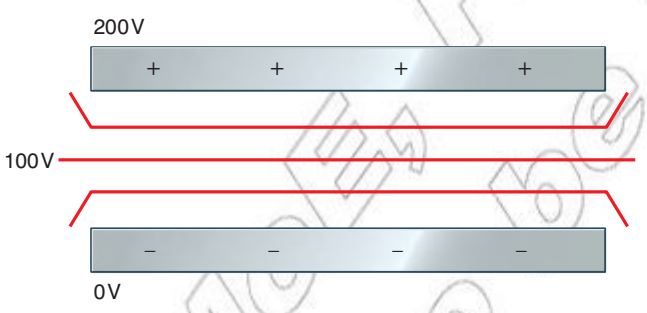


Figure 4.37 *The lines of equipotential between parallel plates are parallel lines. However, these lines curve as the field weakens as you move outside the plates.*

As you might expect the potential reduces steadily from 200 V to 0 V. The potential exactly half way between the plates is 100 V. Any charge placed on this line would have an electrical potential of 100 V.

If the top plate was made more positive then the lines of equipotential would move closer together, but would remain equidistant.

© Ethiopia FDR
Not to be republished

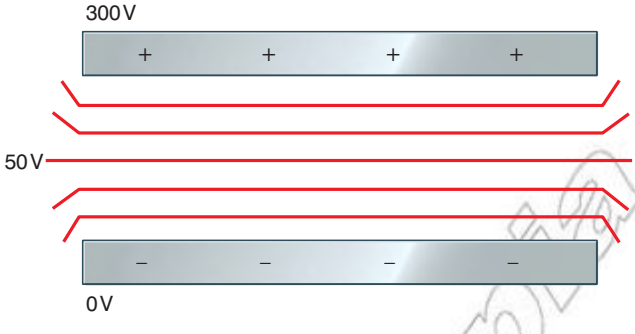


Figure 4.38 *If the electric field strength is increased then the lines of equipotential move closer together.*

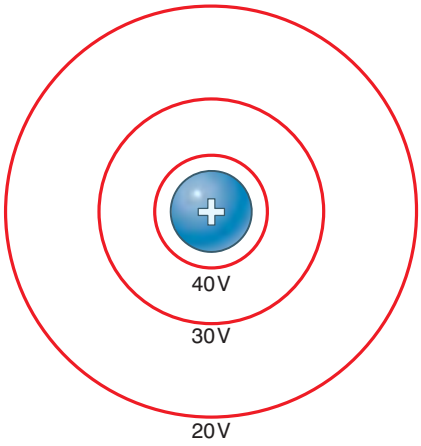


Figure 4.39 *In a radial field the lines of equipotential are further apart as the distance from the charge increases.*

Radial field

In a radial field, such as the field around a point charge, the lines of equipotential are concentric circles centred around the single charge.

These lines move further apart as the distance from the charge increases. This is due to fact the field gets weaker as the distance increases (again from Coulomb's law). As a result, the repulsive force gets weaker and weaker. In order to do the same amount of work you have to move through a greater distance.

Activity 4.12: Equipotential lines

Sketch the lines of equipotential around a –10 C charge and compare these to the lines of equipotential around a 5 C charge.

More complex fields

If multiple charges or different shapes are involved the lines of equipotential can get much more complex. However, in each case they are always perpendicular to the electric field lines.

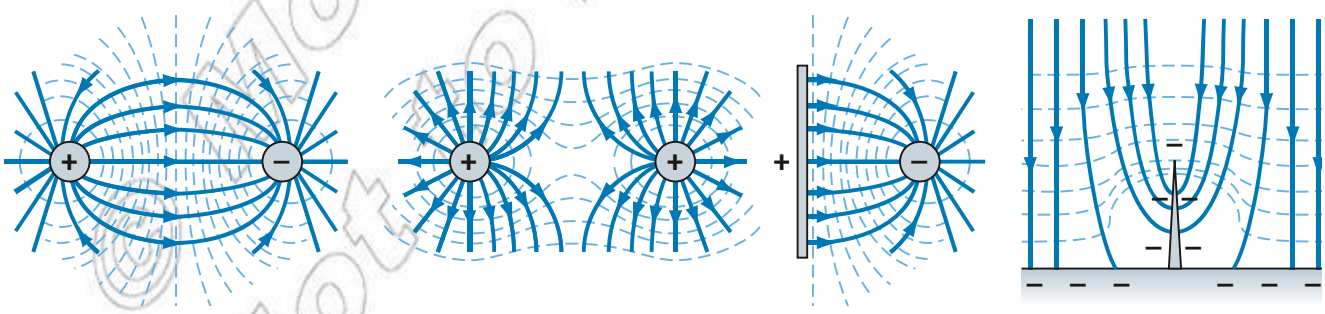


Figure 4.40 *Complex fields still have their equipotentials perpendicular to the field lines at all points.*

© MoE Ethiopia Not to be republished

Measuring equipotentials

Several methods can be used to investigate electric fields. Conducting paper gives a method of obtaining equipotentials and hence field lines for electric fields in two dimensions and for a wide range of electrode shapes. Figure 4.41 shows the setup. The point of a needle connected to a voltmeter is moved over the surface of a sheet of conducting paper, keeping the reading on the voltmeter constant. Using moderate pressure, the path of the needle is traced out on the white paper by means of carbon paper underneath the conducting paper.

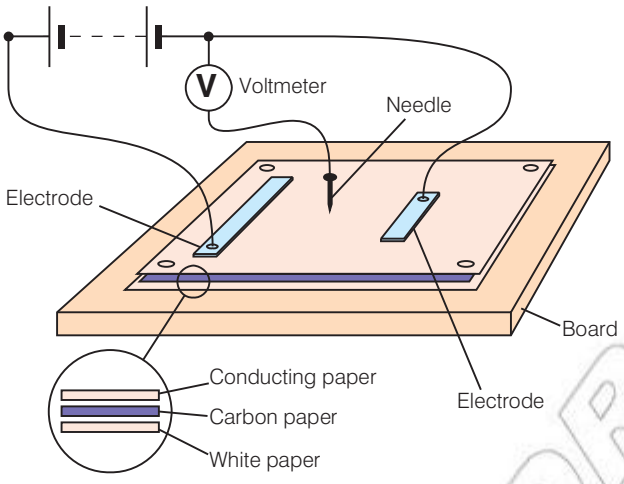


Figure 4.41 *Measuring equipotentials using conducting paper. The voltmeter must have a high resistance so that the potential at a point is not affected by connecting the voltmeter into the circuit.*

What's the relationship between electric field strength and electrical potential?

These two factors are very closely linked. Consider a uniform electric field between two oppositely charged parallel plates.

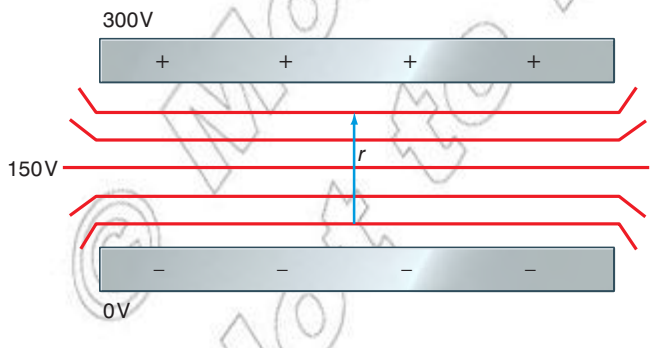


Figure 4.42 *A uniform field between two parallel plates*

To move a charge along line r work would need to be done. This work may be calculated using either:

- $W = Vq$ or $W = Fr$

© MoE FDR Ethiopia Not to be republished

KEY WORDS

potential gradient *the rate of change of electric potential with respect to distance, also equal to the electric field strength*

We can apply the second equation here as it is a uniform field, therefore the force remains constant. Rearranging this gives us:

- $\frac{F}{q} = \frac{V}{r}$

F/q is equal to the electric field strength. Therefore it follows that V/r is also equal to the electric field strength. V/r is referred to as the **potential gradient**. It is more correctly expressed as:

- $E = \frac{\Delta V}{\Delta r}$

The greater the potential gradient, the greater the electric field strength. Moving the same distance in a weaker field will give rise to a smaller change in potential, as seen in Figure 4.43.

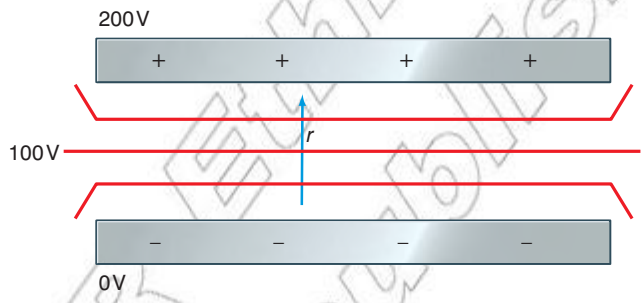


Figure 4.43 *If the potential gradient is shallower (i.e. $\frac{\Delta V}{\Delta r}$ is less) then the electric field strength is also less.*

Imagine plotting a graph of potential against distance moved from the plate (d). For a uniform field it would look like the Figure 4.44.

The gradient of this line is the potential gradient and this is equal to the electric field strength.

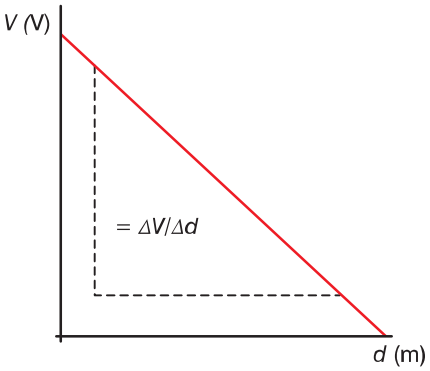


Figure 4.44 *The variation of electric potential with distance from a positively charged parallel plate*

Activity 4.13: Potential gradients

Sketch two more simple graphs of V against d, one showing the plates closer together, the other showing a higher potential on the positive plate. Look carefully at your graphs. In both cases what does the gradient tell you about the electric field strength?

The same is true if you plot a graph of potential against distance form a point positive charge. It will produce a graph similar to the Figure 4.45.

The electric field strength at any distance is given by the gradient of the line at that distance

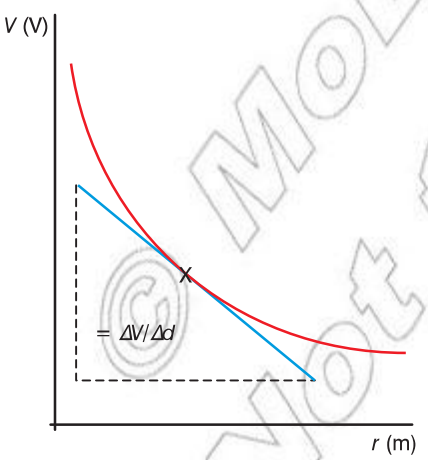


Figure 4.45 *In order to determine the electric field strength at a given distance in a radial field a tangent must be taken at this distance.*

© MoE Ethiopia Not to be republished

Activity 4.14: Potential gradient of a proton

Using $V = \frac{Q}{4\pi\varepsilon_0 r}$ plot a graph of potential against distance for the field around a proton from 1 cm to 10 cm. Determine the electric field strength at 2 cm and 8 cm using the gradient of the line. Confirm the values using $E = \frac{Q}{4\pi\varepsilon_0 r^2}$.

Electrical potential energy

The electrical potential energy of a system can be thought of as the energy of the system due to the particular configuration of the charges within the system.

The electrical potential energy at any point within an electric field is given by

- $E_{EPE} = Vq$

In a radial field around a charge Q this may be written as:

- $E_{EPE} = \frac{Q}{4\pi\varepsilon_0 r} q$ or $E_{EPE} = \frac{1}{4\pi\varepsilon_0} \frac{Qq}{r}$

It is much more common to discuss changes in electrical potential energy, rather than absolute values. As a result it is more helpful to consider.

- $\Delta E_{EPE} = \Delta Vq$

For example, take the case of a uniform field between two oppositely charged parallel plates.

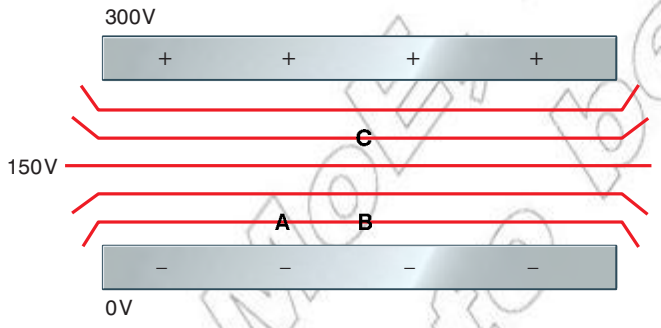


Figure 4.46 *Different positions within an electric field have different potentials. If moving a charge from one position to another with a different electrical potential then the charge will either lose of gain E_{EPE}.*

If a proton moves from A to B there is no change in potential and so there is no change in electrical potential energy. However, if the proton moves from B to C the potential changes from 50 V to 200 V, there is a change is potential of 150 V. The change in E_{EPE} of the proton would be given by:

- $\Delta E_{EPE} = \Delta Vq$ State the relationship
- $\Delta E_{EPE} = 150 \times 1.6 \times 10^{-19}$ Substitute known values
- $\Delta E_{EPE} = 2.4 \times 10^{-17}$ J. Solve equation and give units

The same process may be used in a radial field.

Think about this...

What would the change in E_{EPE} of an electron be if it too moved form B to C? Hint: Think about the work done. Would the electron be gaining or losing E_{EPE}?

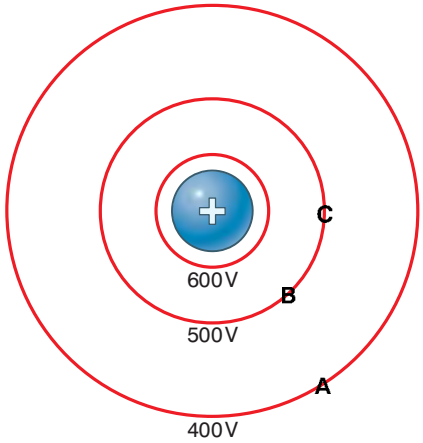


Figure 4.47 *As in uniform fields different positions within a radial electric field have different potentials.*

© MoE FDR Ethiopia Not to be republished

The change in potential from A to B would be 100 V (from B to A it would be –100 V).

We are able to calculate the potential at different points and so determine the change in potential between any two positions (in this case still labelled A and B).

- $\Delta V = \frac{Q}{4\pi\varepsilon_0 r_B} - \frac{Q}{4\pi\varepsilon_0 r_A}$ or $\Delta V = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{r_B} - \frac{1}{r_A}\right)$

Worked example 4.5

Calculate the change in electrical potential energy of a proton as it moves from 5 cm to 0.1 cm from a charged sphere with a charge of 1 nC.

- $\Delta V = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{r_B} - \frac{1}{r_B}\right)$ *State relationship to be used, first find ΔV*
- $\Delta V = \frac{1.0 \times 10^{-9}}{4\pi \times 8.85 \times 10^{-12}}\left(\frac{1}{0.001} - \frac{1}{0.05}\right)$ *Substitute known values*
- ΔV = 8.8 kV Solve for ΔV
- $\Delta E_{EPE} = \Delta V q$ *Use the relationship between ΔE_{EPE} and ΔV*
- $\Delta E_{EPE} = 8800 \times 1.6 \times 10^{-19}$ *Substitute known values*
- $\Delta E_{EPE} = 1.4 \times 10^{-15}$ J. *Solve for ΔE_{EPE}*

If we consider a charged particle moving towards another charged object from a large distance away it is fair to say that the initial potential is zero. This gives us:

- $\Delta V = \frac{Q}{4\pi\varepsilon_0 r} - 0$

For example, consider a proton approaching the very centre of a gold nucleus. If the proton approached from anything more than a few cm then its original potential is negligible. The change in potential to a given point is simply equal to the potential at that point.

As the proton approached the gold nucleus its kinetic energy is converted into E_{EPE}. The proton gets closer and closer to the gold nucleus and as it does so it also gets slower and slower. Eventually it stops before being electrostatically repelled. Applying the law of conservation of energy gives us:

- $\Delta E_k = \Delta E_{EPE}$

This may be expanded to

- $\frac{1}{2}mv^2 = \Delta V q = \frac{Q}{4\pi\varepsilon_0 r} q$

Activity 4.15: Electrical potential energy

Calculate the change in electrical potential energy of an electron as it moves from 10 cm to 2 cm from a charged sphere with a charge of 0.05 mC.

© Not to be republished

Comparing gravitational and electric fields

Section 4.1 has introduced a range of ideas about gravitational and electric fields. We have seen that there are many similarities in the relationships involved in both types of field, and these are summarised in Table 4.1.

We should also remember that there are significant differences between these two types of field too. In particular, we have seen that the comparative strengths of the two types of field are very different, with electric fields being by far the stronger and by far the most important in our lives as far as friction, contact forces and so on are concerned. Also, since mass is always positive, gravitational force is only attractive and gravitational potential always zero or negative. By contrast, since the product Qq may be positive or negative depending on the signs of Q and q, the electrostatic force may be either attractive or repulsive, and electric potential may be zero, positive or negative. This has important consequences in the practical application of these fields – while it is possible to protect, say, delicate electronic components from electric fields using a shield of conducting material, it is impossible to do the same for gravitational fields.

Quantity	Gravitational field	Electric field
Magnitude of force at distance r	$F = \frac{GMm}{r^2}$ (Force is always attractive)	$F = \frac{1}{4\pi\varepsilon_0}\frac{\lvert Qq\rvert}{r^2}$ (Force may be attractive or repulsive)
Magnitude of field strength at distance r	$g = \frac{F}{m} = \frac{GM}{r^2} = \frac{dV}{dr}$ (Field is always radially *in*, potential gradient always radially *out*)	$E = \frac{F}{\lvert q\rvert} = \frac{1}{4\pi\varepsilon_0}\frac{\lvert Qq\rvert}{r^2} = \left\lvert\frac{dV}{dr}\right\rvert$ (For negative charge, field is radially *in* and potential gradient is radially *out*, and vice versa for positive charge)
Potential energy at distance r	$E_p = -\frac{GMm}{r}$	$E_p = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r}$
Potential at distance r	$V = -\frac{GM}{r}$	$V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r}$

Table 4.1 *Comparing gravitational and electric fields. The vertical lines, e.g. $|q|$, mean 'take the magnitude of'. This is necessary since we are concerned here with the magnitude of the force or field strength, which is always positive.*

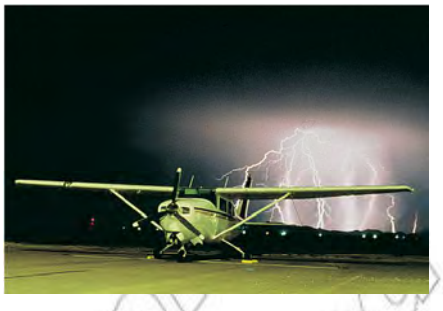


Figure 4.48 *Inside a hollow conductor, the electric field is zero even when there is a very strong electric field outside the conductor. The metal body of this plane forms a good shield against atmospheric electric fields, so that the occupants are quite safe from atmospheric electrical discharges (lightning!). Hollow conducting shields like this are called Faraday cages. They are effective because of the two types of electric charge – there is no equivalent of the Faraday cage for gravitational fields.*

Summary

In this section you have learnt that:

- Electric potential is defined as the work done per unit positive charge to move a positive test charge from infinity its current position within an electric field. It is a scalar quantity with units of V or J/C.
- The electric potential around a point charge is given by $V = \frac{Q}{4\pi\varepsilon_0 r}$ or $V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r}$.
- An equipotential is a line joining points within an electric field with the same potential. In a uniform field the lines of equipotential are equidistant parallel lines. In a radial field, such as the field around a point charge, the lines of equipotential are concentric circles centred around the single charge (these lines get further apart as the distance from the charge increases).
- The electric field strength at a point is equal to the potential gradient at that point $E = \frac{\Delta V}{\Delta r}$.
- The change in electrical potential is equal to ΔVq. In a radial field this may be written as $\Delta E_{EPE} = \frac{Q}{4\pi\varepsilon_0 \Delta r} q$.

Review questions

1. Define electrical potential.
2. Calculate the electrical potential 3 mm from of point charge of 5.0 nC.
3. Draw the field lines and the lines of equipotential:
   a) between two oppositely charged parallel plates
   b) around a negative point charge.
4. Calculate the change in electrical potential energy when an electron moves towards a proton from an initial distance of 0.1 mm to a distance of 0.1 μm.
5. An alpha particle with kinetic energy 5.1×10^{-13} J is fired at a uranium nucleus. Calculate how close the alpha particle gets to the uranium nucleus. The charge on the alpha particle is $+2e$, and that on the uranium nucleus is $+92e$. (Assume that the uranium nucleus remains stationary throughout, and treat both the alpha particle and the uranium nucleus as spheres of charge.)

© MoE, Not to be republished

4.3 Capacitors and dielectrics

By the end of this section you should be able to:

- Derive the formula for a parallel plate capacitor (from Gauss' law), including use of a dielectric.
- Define the dielectric constant.
- Explain qualitatively the charge and discharge of a capacitor in series with a resistor.
- Explain the behaviour of an insulator in an electric field.
- Define electric energy density and derive the formula for the energy density for an electric field using a parallel plate capacitor.
- Solve problems involving capacitances, dielectrics and energy stored in a capacitor.

Capacitors and capacitance

An electric field can cause charged particles to move. Indeed, this is why a current flows through a circuit – an electric field is set up within the conducting material and this causes electrons to feel a force and thus move through the wires and components of the circuit. Where there is a gap in a circuit, although the effect of the electric field can be felt by charges across the empty space, conduction electrons are generally unable to escape their conductor and move across the gap. This is why a complete path is needed for a simple electric circuit to function.

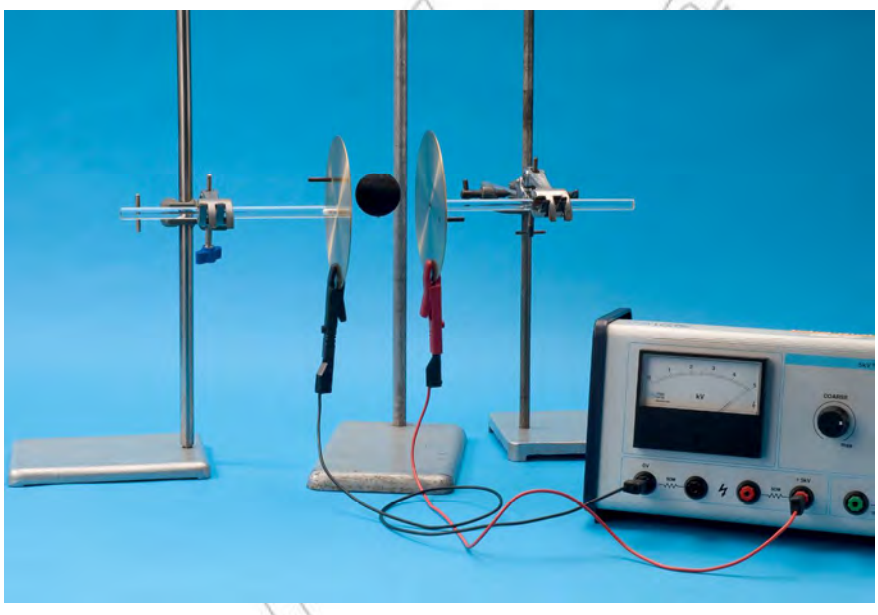


Figure 4.49 *An electric field acts across a space. You could test this by hanging a charged sphere near the plates and observing the field's force acting on the sphere.*

Ethiopia Not to be republished

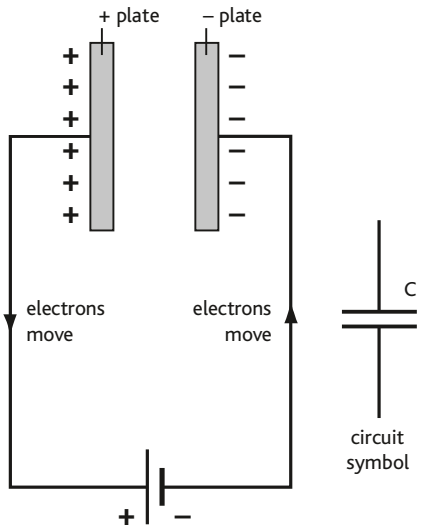


Figure 4.50 *A simple capacitor circuit.*

DID YOU KNOW?

One Farad is a very large capacitance! Most capacitors have capacitances in the range of mF to pF.

Think about this...

Take an example of a charged capacitor with –6 mC on the negative plate. This capacitor will have a charge of +6 mC on the positive plate. This can lead to erroneous statements regarding the charge stored. It is often mistaken for 0 mC or even 12 mC. The capacitor in this example currently stores 6 mC of charge (this is the charge that will flow around the circuit when the capacitor is discharged.

KEY WORDS

capacitor *an electrical device characterised by its capacity to store an electric charge*
capacitance *an electrical phenomenon whereby an electric charge is stored*

However, charge can be made to flow in an incomplete circuit. This can be demonstrated by connecting two large metal plates in a circuit with an air gap between them (Figure 4.49). The circuit shown in Figure 4.50 represents the situation shown by the photo in Figure 4.49. When the power supply is connected, the electric field created in the conducting wires causes electrons to flow towards the positive terminal. Since the electrons cannot cross the gap between the plates they build up on the plate connected to the negative terminal, which becomes negatively charged. Electrons in the plate connected to the positive terminal flow towards the positive terminal, resulting in a positive charge on that plate. The attraction between the opposite charges across the gap creates an electric field between the plates which increases until the pd across the plates is equal to the p.d. of the power supply.

A pair of plates like this with an insulator between them is called a **capacitor**. As we have seen, charge will build up on a capacitor until the pd across the plates equals that provided by the power supply to which it is connected. At that stage it is said to be fully charged. The capacitor is acting as a store of charge. The amount of charge a capacitor can store, per volt applied across it, is called its **capacitance**, C, and is measured in farads (F). The capacitance depends on the size of the plates, their separation, and the nature of the insulator between them.

Capacitance can be calculated from the equation:

$$C = \frac{Q}{V}$$

Worked example 4.6

a What is the capacitance of a capacitor which can store 18 mC of charge when the pd across it is 6 V?

$$C = \frac{Q}{V}$$

$$= \frac{18 \times 10^{-3}}{6}$$

$$= 3 \times 10^{-3}$$

$$C = 3\,\text{mF}$$

b How much charge will be stored on this capacitor if the voltage is increased to 20 V?

$$Q = CV$$

$$= 3 \times 10^{-3} \times 20$$

$$= 60 \times 10^{-3}$$

$$= 0.06\,\text{C}$$

Activity 4.16: Capacitances

Complete the table below:

Q (C)	*V* (V)	*C* (F)
1.2 μ	6.0	
2.5×10^{-6}		1000 p
	120	3.0×10^{-8}

The greater the p.d. across the capacitor the greater the amount of charge it can store. This may be seen in the graph below.

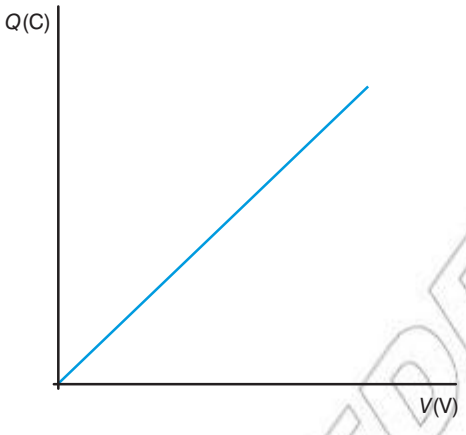


Figure 4.52 *$Q \propto V$ for a capacitor. When plotting a graph of Q against V the gradient of the line is equal to the capacitance.*

However, this can not continue to increase indefinitely. Eventually the p.d. across the plates will become too high. Charge will begin to spark across from one plate to the other. When this happens the capacitor is said to be **breaking down**. It ceases to store charge and begins to conduct electricity.

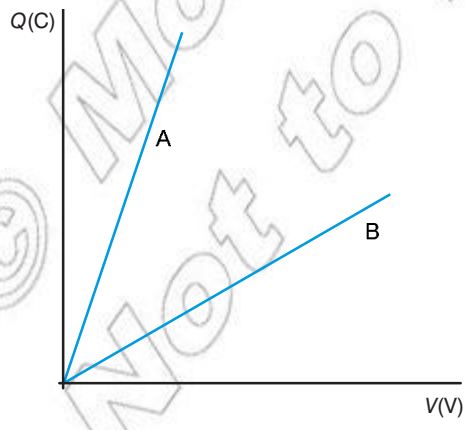


Figure 4.53 *Two different capacitors A and B; in this example, capacitor A has a higher capacitance than capacitor B.*

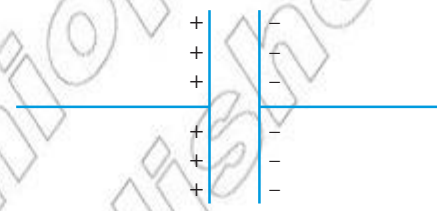


Figure 4.51 *Due to electrostatic attraction the charge remains on the plates even when the supply is disconnected. Some capacitors can retain their charge for several months.*

KEY WORDS

breaking down *when the voltage applied across a capacitor is too high, the dielectric ceases to act as an insulator and the charge starts to spark across the plates*

© MoE FDR Ethiopia Not to be republished

Gauss' law and capacitance

In section 4.1 we used Gauss's law to show the field between two parallel plates is given by:

- $E = \frac{\sigma}{\varepsilon_0}$

where σ was the charge per unit area on each plate. Therefore this relationship can be written as:

- $E = \frac{\sigma}{\varepsilon_0} = \frac{Q}{A\varepsilon_0}$

The potential difference between each plate may be given by:

- $V = Ed$

Therefore

- $V = (Q / A\varepsilon_0)\, d$
- $V = \frac{Qd}{A\varepsilon_0}$

Substituting this in to our defining equation for capacitance we get:

- $C = \frac{Q}{V}$
- $C = Q / (Qd / A\varepsilon_0)$

This cancels to give:

- $C = \frac{\varepsilon_0 A}{d}$

This equation gives the capacitance of two parallel plates in a vacuum.

KEY WORDS

dielectric *an insulating material that may be polarised by an electrical field and which allows more charge to be stored*

dielectric constant *the ratio of the permittivity of the dielectric material to the permittivity of free space*

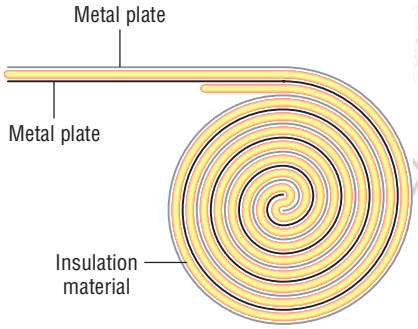


Figure 4.54 *This diagram shows how the plates of a capacitor are rolled up on one another to make the component more compact for fitting into electrical circuits*

How are practical capacitors constructed?

In reality simple parallel plate capacitors have very little real use as they are limited to capacitances of around 10^{-14} F. If you continued to increase the p.d. across the capacitor it would simply break down.

Most capacitors are constructed with a **dielectric** material in between the plates rather than just an air gap. This is an insulating material with properties that allow more charge to be stored.

The capacitance of a given capacitor is given by:

- $C = \frac{\varepsilon_0 \varepsilon_r A}{d}$

where

A = surface area of plate

d = distance between plates

ε_r = relative permittivity of the material between the plates (also called the **dielectric constant**). This is the permittivity of the dielectric relative to the permittivity of free space. For example, an ε_r of 2 would mean double the permittivity (it is twice as 'easy' for the electric field to travel through the space between the plates).

© FDRE Ethiopia to be republished

In order to make a capacitor as large as possible we could do the following.

Make the area of the plates as large as possible

This is often accomplished by rolling two metal plates around each other with an insulator in between. This has the effect of dramatically increasing the area of the plates but with relatively little increase in capacitor volume.

Move the plates as close together as possible

However, the closer the plates are together, the lower the breakdown voltage. Apply too high a p.d. and the capacitor will become conducting.

Use a dielectric between the plates with as large a dielectric constant as possible

A dielectric is an electrical insulator that may be polarised by an electric field. This has the effect of dramatically increasing the charge stored at a given p.d.

A dielectric contains a series of **dipoles** (or molecules that will become dipoles when a field is applied). In this case the dipole is just simply a molecule with positive and negative ends. These dipoles are usually randomly organised.

As we've already seen in section 4.1, when an electric field is applied to the dielectric charges do not flow through the material (like they would in a conductor), but they do cause the dipoles to rotate and line up with the electric field.

The use of a dielectric dramatically increases the permittivity of the region in between the plates and so allows much more charge to be stored at the same p.d. Table 4.2 shows the relative permittivity for a number of different materials.

***Table 4.2** Relative permittivity of different materials*

Material	ε_r (no units as relative to ε_0)
Vacuum	1 – by definition
Perspex	3.3
Mica	7
Water	80
Barium titanate	1200

As discussed a dielectric constant of 7 means the field strength between the plates would be 7 times greater than if there was a just vacuum between the plates.

Each material also has a **dielectric strength**, this is maximum electric field strength that it can withstand without breaking down. Above this field strength the dielectric will break down and begin to conduct. This causes its capacitance to fall to zero.

KEY WORDS

dipoles *a pair of electric charges or magnetic poles, of equal magnitude but of opposite sign or polarity, separated by a small distance*
dielectric strength *the maximum electric field strength that a material can withstand before breaking down*

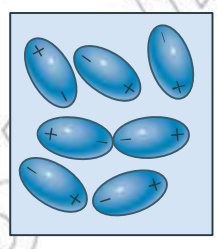


***Figure 4.55** Dipoles in a dielectric are usually randomly arranged.*

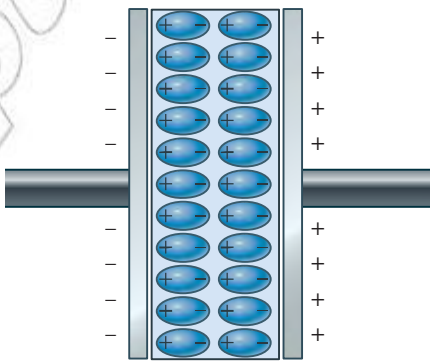


***Figure 4.56** With the application of an electric field the dipoles all line up and so increase the capacitance of the capacitor.*

DID YOU KNOW?

Electrolytic capacitors contain an ionic conducting liquid as one of its plates; this allows for even more charge to be stored at the same p.d. These capacitors must be connected to the correct polarity. If they are wired up the wrong way the liquid will rapidly heat up. This often causes the capacitor to burst or explode.

Not for Distribution

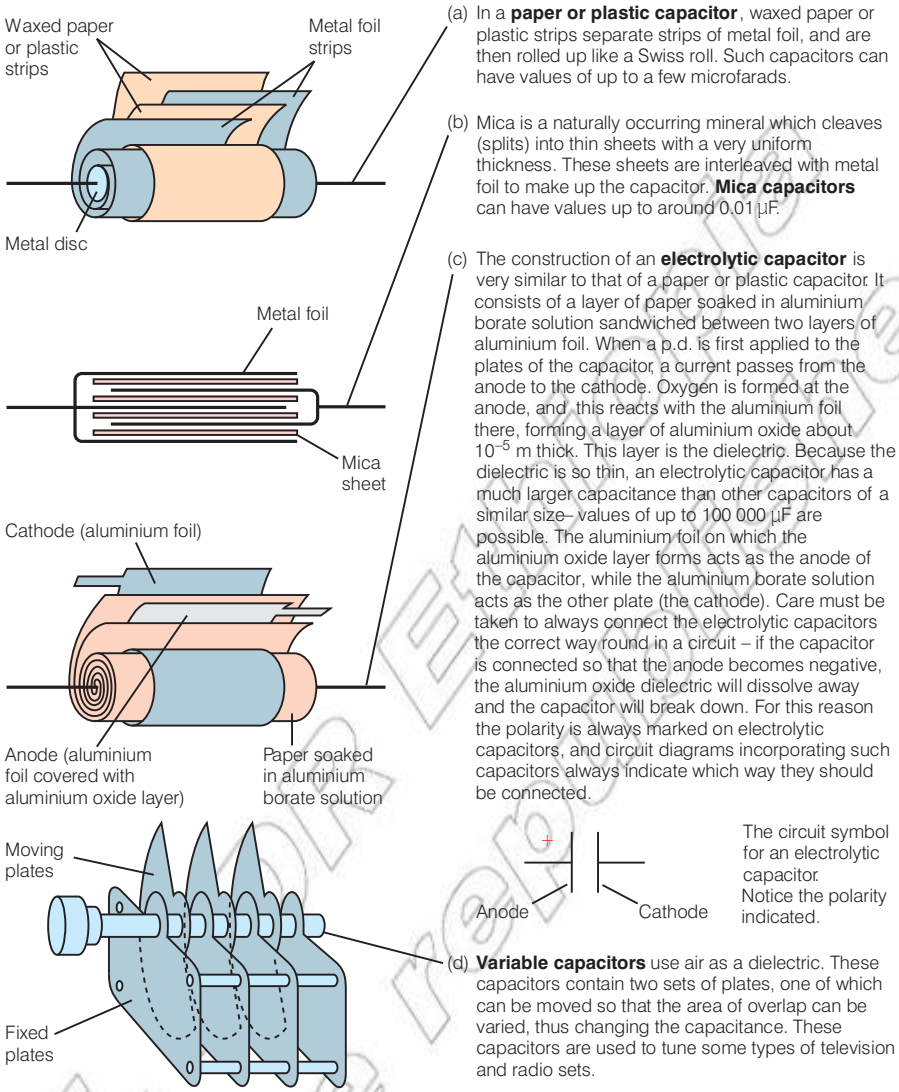


***Figure 4.57** Different types of capacitors*

The Leyden jar

The Leyden jar (sometimes called Leiden jar) is an example of an early device used to store charge. It is the forerunner to modern capacitors.

The Leyden jar played a key role in several important early electrostatics experiments.

© MoE, Not to be republished

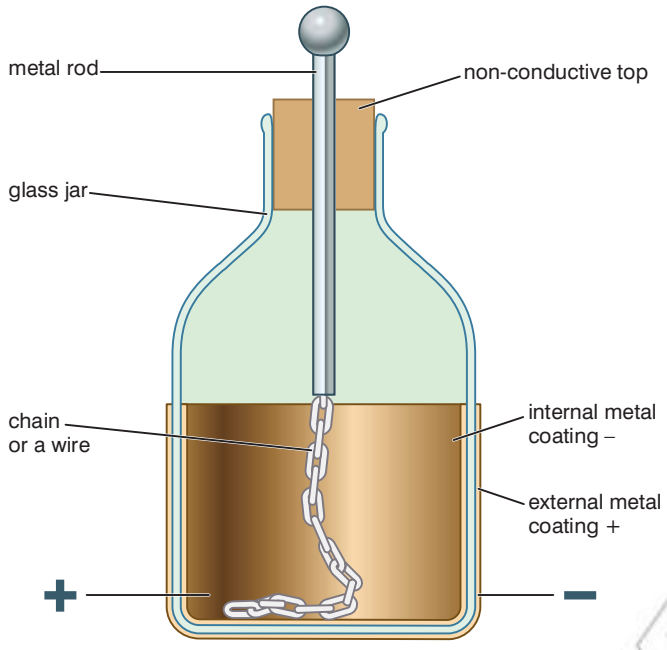


Figure 4.58 *An example of a simple Leyden jar*

Most Leyden jars have similar designs. They are constructed using a glass jar with foil covering the inner and outer surfaces of the jar. The inner surface is usually connected to a rod with may be electrostatically charged. The inner surface then develops the same charge and as the outer foil is earthed the two surfaces behave like oppositely charged parallel plates.

What happens when you connect capacitors in series and parallel?

Capacitors in parallel

In a circuit, capacitors can often be in parallel with one another. One such circuit is shown in Figure 4.59. Here three capacitors, C_1, C_2 and C_3, are connected in parallel with one another and have a battery connected to them to provide a potential difference V across each of them. The process of charging each capacitor will be just the same as if they were connected individually to the battery. This means that the charge Q_1 on C_1 will be C_1V, etc. This will give the total charge Q on all the capacitors as

$$Q = C_1V + C_2V + C_3V = (C_1 + C_1 + C_3)\text{V}$$

The total capacitance C of all the capacitors is the total charge/p.d. across them, so

$$C = \frac{Q}{V} = C_1 + C_2 + C_3$$

Note that it is when capacitors are in parallel that their individual capacitances can be added directly to find the total capacitance. This has to be so. For a given p.d., more charge is being stored, and so the capacitance must be greater. The above equation can be extended for any number of capacitors in parallel.

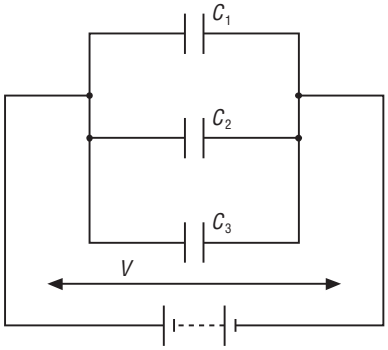


Figure 4.59 *Three capacitors in parallel with the same potential difference V across each one*

© MoE-EDP Ethiopia Not to be republished

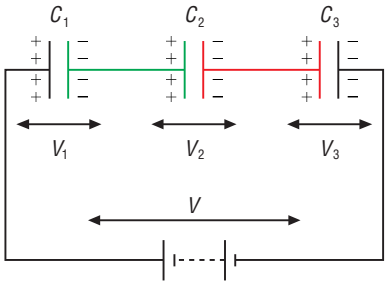


Figure 4.60 *Three capacitors in series. Each has the same charge*

Capacitors in series

When three capacitors are in series with one another, the circuit will be as shown in Figure 4.60.

The first point to note about the circuit is that the application of Kirchhoff's second law means that

$V = V_1 + V_2 + V_3$

To understand the second point, look at the two plates on adjacent capacitors and the wire between them coloured green. These plates and the wire are not actually connected to the battery. As they are uncharged before the battery is connected, they must be uncharged (in total) after connection. Any positive charge on one plate must mean an equal amount of negative charge on the other plate. The same is true for the two plates and connecting wire for the plates coloured red. This is why the charges have been marked on the plates. Every plate has the same magnitude charge on it in a series circuit. The charge Q, supplied from the battery, is equal to the charge on all three capacitors. This instinctively sounds wrong; it seems as though Kirchhoff's first law must be broken when, for instance, 4 mC leaves the battery to charge up three capacitors, each capacitor has 4 mC on it. But the total charge is 4 mC, *not* 12 mC. This now enables the total capacitance of the circuit to be found.

Since

$V = V_1 + V_2 + V_3$

$\frac{Q}{C} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3}$ and since Q cancels out throughout, we get

$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$

As before, this expression can be extended for as many capacitors as there are in series.

© MoE: Namibia Not to be reproduced

Worked example 4.7

A capacitor of capacitance 0.00100 μF has a 12.0 V battery connected across it, as shown in Figure 4.61(a).

a) Calculate the charge on the capacitor.

b) A break develops in the circuit at A. The two ends of the wire at the break are near to one another, so they behave as a capacitor of capacitance 20 pF. The circuit effectively becomes the circuit in Figure 4.61(b). When this broken circuit is switched on, with both capacitors initially uncharged, what will be:

(i) the total circuit capacitance?
(ii) the charge on each capacitor?
(iii) the p.d. across each capacitor?

Answer

a) $Q = CV = 0.00100\ \mu F \times 12\ V = 0.012\ \mu C$ (Note how the μ symbol can be carried through the equation.)

b) (i) $$\frac{1}{C} = \frac{1}{0.00100} + \frac{1}{0.000020} = 1000 + 50000 = 51000$$

$$C = \frac{1}{51000} = 1.96 \times 10^{-5}\ \mu F$$

(ii) The charge Q on each capacitor and the total charge are the same, so

$$Q = CV = 12\ V \times 1.96 \times 10^{-5}\ \mu F = 2.35 \times 10^{-4}\ \mu C$$

(iii) The p.d. across the 0.00100 μF capacitor is

$$\frac{Q}{C} = \frac{2.35 \times 10^{-4}\ \mu C}{0.00100\ \mu F} = 0.24\ V$$

The p.d. across the 0.000020 μF capacitor is

$$\frac{Q}{C} = \frac{2.35 \times 10^{-4}\ \mu C}{0.000020\ \mu F} = 11.76\ V$$

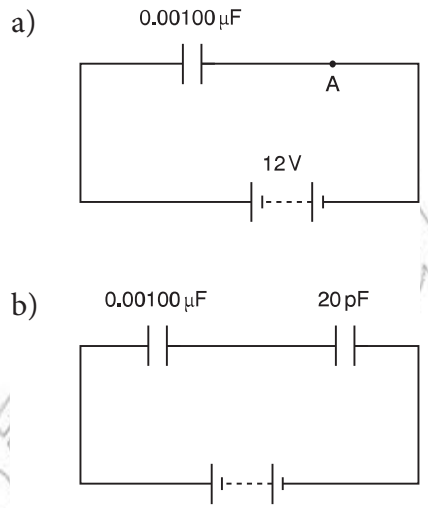


Figure 4.61 *(a) A single capacitor in a circuit; (b) the modified circuit after a fault develops*

The worked example shows again how careful you will need to be with powers of 10. When dealing with numbers less than one, far more mistakes can be made than with numbers greater than one.

Note that in the worked example the large p.d. appears across the small capacitor. An electrician may try to find a fault in a circuit by connecting a voltmeter across various components. The components across which there is zero p.d. are probably working well. If there is a high p.d. across a resistor, then that resistor is probably the one in which there is a break. For the same reason there will normally be a high p.d. across a switch that is off; when it is switched on, the p.d. across it will be zero.

© Mindset Learn Not to publish

How much energy can a capacitor store?

Although capacitors are incredibly useful in electronic devices they only store a relatively small amount of energy and certainly not enough to be used in large-scale electricity storage and power distribution.

If we consider a graph of p.d. across a capacitor against the charge stored, assuming it is below the breakdown voltage, it will appear as in Figure 4.62

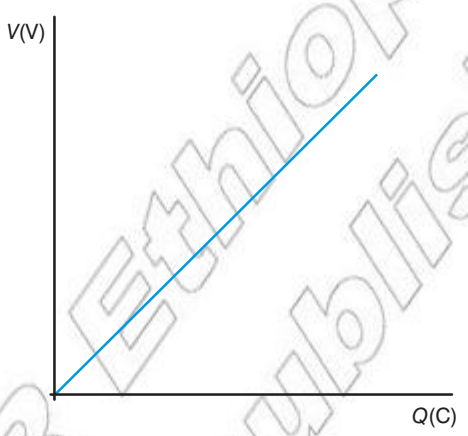


Figure 4.62 *When considering the energy stored by a capacitor it is helpful to consider a graph of p.d. against charged stored. In this case the gradient of the line is equal to 1/C.*

We have already seen this relationship; however, notice in this case we are plotting V against Q not Q against V. Both graphs are often used, so ensure you study them carefully before answering any questions.

If we consider a small region under the line the p.d. will remain constant. The area of the rectangle will be equal to $V\delta Q$. This is the small increase in the energy stored by the capacitor.

Therefore the total area under this line is equal to the total energy stored in the capacitor.

Think about this...

In order to charge a capacitor a supply (cell, battery or power supply) does work equal to QV. Yet the capacitor only stores $\frac{1}{2}QV$; where does the remaining energy go? Hint: think about the connecting wires.

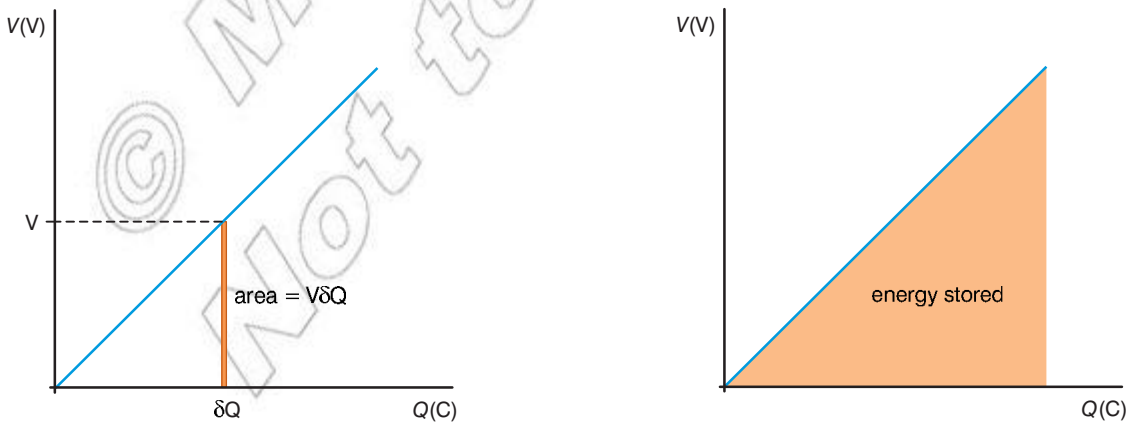


Figure 4.63 *Calculus may be used to determine the total energy stored. However, as the graph is a straight line it is simple to see the total area under the line is equal to $\frac{1}{2}QV$.*

© MoE-EPDR Ethiopia Not to be republished

The energy stored by a capacitor is therefore given by:

- $E = \frac{1}{2}QV$

This equation may be combined with $Q = VC$ to give:

- $E = \frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{Q^2}{2C}$

Activity 4.17: Energy storage

Complete the table of missing values below:

Q (C)	V (V)	C (F)	E (J)
5.0×10^{-3}	12		
	200	3.0×10^{-6}	
20×10^{-6}			6.0×10^{-3}

DID YOU KNOW?

Even a large capacitor does not really store that much energy. Take a 10 mF capacitor (a relatively large capacitor) charged to 200 V (a high p.d. for such a large capacitor). This would only store 1 J! Enough to lift an apple 1 m into the air.

Electric energy density

KEY WORDS

electrical energy density *the electrical energy stored per unit volume of a capacitor*

The **electrical energy density** (u) of a capacitor is simply the electrical energy stored per unit volume.

- u = energy stored / volume of capacitor

The volume of the region between the parallel plates is given by:

- volume = Ad

where

A = area of each plate

d = distance between plates

Therefore the electrical energy density may be found by:

- u = energy stored / volume of capacitor
- $u = \frac{\frac{1}{2}CV^2}{Ad}$

However:

- $C = \frac{\varepsilon_0 \varepsilon_r A}{d}$

Therefore substituting this in we get:

- $u = \frac{1}{2}(\varepsilon_0 \varepsilon_r A / d) V^2 / Ad$

This simplifies to:

- $u = \frac{\frac{1}{2}\varepsilon_0 \varepsilon_r V^2}{d^2}$

However $E = V / d$ therefore $E^2 = V^2 / d^2$, giving:

- $u = \frac{1}{2}\varepsilon_0 \varepsilon_r E^2$

u = electrical energy density of the capacitor in J/m

For any given parallel plate capacitor the electrical energy density is proportional to the dielectric constant and the square of the electric field strength between the plates.

© MoE: FDR Ethiopia Not to be republish

Discharging a capacitor

We can investigate the discharge of a capacitor using the circuit shown in Figure 4.64.

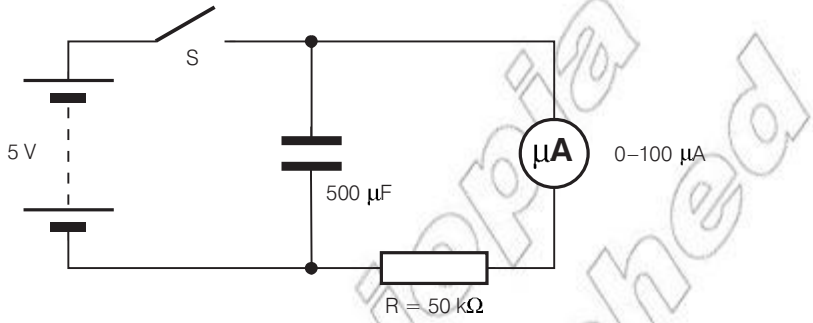


Figure 4.64

The capacitor is charged by closing the switch S. When the switch is closed, the microammeter reads 100 μA (I = 5 V/50 kW = 10^{-4} A). At the instant S is opened, this reading begins to fall.

The p.d. across the capacitor, the charge on it and the current through the circuit at any instant are related by two equations:

$V = IR$ and $C = Q/V$

We also know that the current through the circuit at any instant is the rate of flow of charge from the capacitor at that instant, so we can write:

$$I = -\frac{\Delta Q}{\Delta t}$$

(The negative sign shows that the charge on the capacitor decreases as time increases.)

The instantaneous current through the circuit is also related to the p.d. across the capacitor and the resistance of the circuit:

$$I = \frac{V}{R}$$

And the p.d. across the capacitor at any instant is related to the charge on it at that instant and its capacitance:

$$V = \frac{Q}{C}$$

Bringing these three equations together:

$$I = -\frac{\Delta Q}{\Delta t} = \frac{Q}{RC}$$

Rearranging this:

$$\Delta Q = -\frac{Q}{RC}\,\Delta t$$

We can now use this relationship to calculate the charge on the capacitor at 5-second intervals as it discharges, and to calculate the current at those times too. Figure 4.65 on page 186 shows how this is done for the example in Figure 4.64.

© MoE: FDRE Ethiopia
Not to be republished

The graphs of current and charge versus time have a constant period in which a quantity (in this case charge or current) halves. This is characteristic of an exponential decay, in which the rate of decrease of a quantity is proportional to the quantity itself.

(Note that the p.d. across the capacitor is directly related to the charge on it, and that the current through the circuit is directly related to the p.d. across the capacitor – therefore all these quantities vary in a similar way.)

Calculus is required to investigate fully the relationships involved in the discharge of a capacitor, but some very simple mathematics together with a graphical treatment using techniques like those used in Figure 4.65 show that:

- The time taken for the capacitor to discharge from voltage V to voltage $V/2$ is proportional to RC (the resistance of the circuit multiplied by the capacitance of the capacitor) – the quantity RC is called the time constant of the circuit.
- The decay of charge on the capacitor has a constant half-life of just over two- thirds of the time constant (actually $0.693RC$).
- After it has been discharging for a time RC, the charge on a capacitor has fallen to a little over one-third of its initial value (actually $0.37Q_0$).
- At first it may seem surprising that the time constant RC is a measure of the time taken for the capacitor to discharge. However, a little thought suggests that this is not unreasonable, since:
- Increasing R decreases the current through the circuit, thus increasing the time the capacitor takes to discharge.
- Increasing C increases the charge on the capacitor for a given p.d. across it, without changing the current through the circuit.

In addition, multiplying resistance by capacitance results in a quantity with the units of time:

$$\Omega \times \text{F} = \text{V/A} \times \text{C/V} = \text{V/(C/s)} \times \text{C/V}$$
$$= \text{s}$$

From considering the discharge of a capacitor we know that we can write:

$$\frac{\Delta Q}{\Delta t} = -\frac{Q}{RC}$$

If we now let $\Delta t \rightarrow 0$, the differential equation which results may be solved by integration:

$$\int_{Q_0}^{Q} \frac{\text{d}Q}{Q} = -\int_{0}^{t} \frac{\text{d}t}{RC}$$

The limits of the integration are chosen so that the charge on the capacitor is Q_0 when $t = 0$ and Q when $t = t$.

© MoE FDRE Ethiopia Not to be republished

Write down p.d. across capacitor at start of investigation in third column of table.

Calculate charge on capaciitor at $t = 0$ using $Q = CV$. Record this in second column of table.

Calculate current through circuit using $I = V/R$. Record this in fourth column of table.

Calculate charge leaving capacitor in the next 5 seconds using $\partial\theta = I\partial t$. Record this in fifth column of table.

Calculate charge left on capacitor at end of this 5-second period. Record this in second column of table on next line.

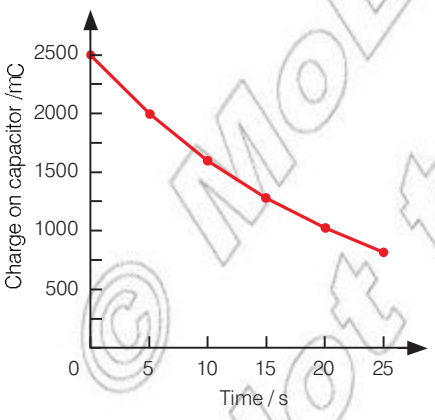
Calculate p.d. across capacitor now using $V = Q/C$. Record this in third column of table.

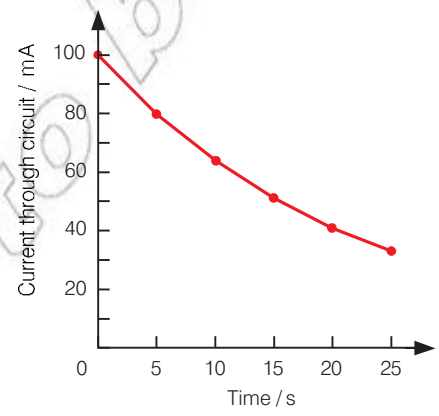
Plot graphs of (1) the charge on the capacitor versus time and (2) the current through the circuit versus time

Repeat to complete table.

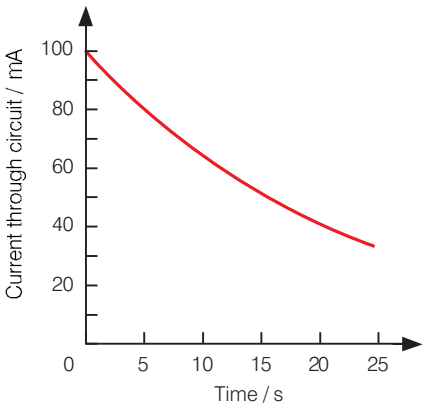
Decay of charge on a capacitor

Time / s	*Charge* / mC	*P.d.* / V	*Current* / mA	DQ / mC
0	2500	5.0	100	500
5	2000	4.0	80	400
10	1600	3.2	64	320
15	1280	2.56	51	256
20	1024	2.05	41	205
25	819	1.64	33	164





These two graphs show the charge on the capacitor and the current through the circuit calculated at 5-s intervals.



This graph shows the current through the circuit as measured constantly in an experiment

Figure 4.65

© MoE FDR Ethi... Not to be republis...

Integrating this relationship gives the result:

$$\left[\log_e Q\right]_{Q0}^{Q} - = - \left[\frac{t}{RC}\right]_0^t$$

When these two expressions are evaluated between their limits, we get:

$\log_e Q - \log_e Q_0 = - t/RC$

Since $\log x - \log y = \log (x/y)$, this becomes:

$$\log_e \frac{Q}{Q_0} = - \frac{t}{RC}$$

or:

$Q = Q_0 e^{-t/RC}$

The time taken for the capacitor to lose half its charge is known as the half-life $t_{1/2}$ of the decay process. In this case, $Q = Q_0/2$ when $t = t_{1/2}$, so:

$$\log_e \frac{Q_0/2}{Q_0} = - \frac{t_{1/2}}{RC}$$

or:

$$\log_e \frac{1}{2} = - \frac{t_{1/2}}{RC}$$

Since $-\log_e \frac{1}{2} = \log_e 2$ this can be rearranged:

$t_{1/2} = RC \log_e$

$= 0.693\ RC$

When t = the time constant, *RC*:

$Q = Q_0 e^{-RC/RC} = Q_0 e^{-1}$

$\approx 0.37 Q_0$

Thus *RC* is the time for the charge on the capacitor to fall to 0.37 times its initial value, as we have already seen.

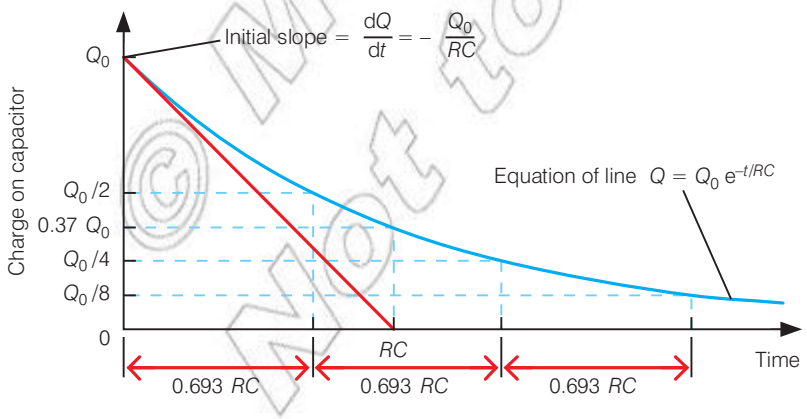


Figure 4.66 *The exponential decay curve of the charge on a capacitor shows a constant half-life of RC $\log_e$ 2. The graph also illustrates how the capacitor would fully discharge in time RC if it continued to discharge at its initial rate, and how the actual charge remaining on it after this time is 0.37Q_0.*

© MoE, FDR Ethiopia Not to be republished

Worked example 4.8

A 50 μF capacitor is discharged through a 10 000 Ω resistance. How long will it take for the potential difference across the capacitor to fall to 40% of its initial value?

We know that for a capacitor C discharging through a resistor R, the charge remaining after time t has elapsed is given by:

$$Q = Q_0 e^{-t/RC}$$

where Q_0 is the charge on the capacitor at $t = 0$.

Since $C = Q/V$, it follows that $V \propto Q$, and we may also write:

$$V = V_0 e^{-t/RC}$$

or:

$$\log_e V = \log_e V_0 - \frac{t}{RC}$$

This can be simplified to:

$$\frac{t}{RC} = \log_e \frac{V_0}{V}$$

In this case, $V = 0.4V_0$, $R = 10\,000\ \Omega$ and $C = 50 \times 10^{-6}$ F, so:

$$\frac{t}{10\,000\ \Omega \times 50 \times 10^{-6}\ \text{F}} = \log_e \frac{V_0}{0.4\ V_0}$$

so:

$$t = (10\,000\ \Omega \times 50 \times 10^{-6}\ \text{F}) \times \log_e 2.5$$
$$= 0.46\ \text{s}$$

The potential difference across the capacitor falls to 40% of its initial value in 0.46 s.

Charging a capacitor

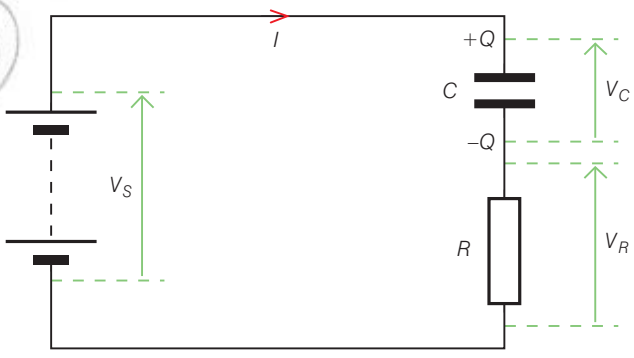


Figure 4.67

© MoE Not to be

A similar process enables the charging of the capacitor to be modelled. In the circuit shown in Figure 4.67, Kirchhoff's loop rule (see section 5.2) tells us that:

$$V_S = V_R + V_C = IR + \frac{Q}{C}$$

As the capacitor charges, I and Q will change while V_S and C will not, so that in a time interval of Δt we can write:

$$\frac{\Delta V_S}{\Delta t} = 0 = \frac{\Delta I\, R}{\Delta t} + \frac{1}{C}\frac{\Delta Q}{\Delta t}$$

But $\Delta Q/\Delta t = I$, so:

$$0 = \frac{\Delta I\, R}{\Delta t} + \frac{I}{C}$$

Rearranging gives us:

$$\frac{\Delta I}{\Delta t} = -\frac{I}{RC}$$

If $\Delta t \rightarrow 0$, this produces a differential equation:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = -\frac{I}{RC}$$

This is exactly the same form of the equation we saw before in the case of the decay of charge on the capacitor. The solution to this equation is also obtained by integration:

$$\int_{I_0}^{I} \frac{\mathrm{d}I}{I} = -\int_{0}^{t} \frac{\mathrm{d}t}{RC}$$

This produces the solution:

$$I = I_0\, \mathrm{e}^{-t/RC}$$

Now when $t = 0$, we know that:

$I = I_0 = V_S/R$, since at $t = 0$, $V_C = 0$

We also know that $I = V_R/R$, so:

$$I = \frac{V_R}{R} = \frac{V_S}{R}\mathrm{e}^{-t/RC}$$

that is,

$$V_R = V_S\, \mathrm{e}^{-t/RC}$$

Now the p.d. across the capacitor is given by:

$$V_C = V_S - V_R$$

so:

$$V_C = V_S - V_S\, \mathrm{e}^{-t/RC}$$

$$= V_S(1 - \mathrm{e}^{-t/RC})$$

$$Q = CV_S(1 - \mathrm{e}^{-t/RC})$$

© MoE, FDR Ethiopia Not to be republished

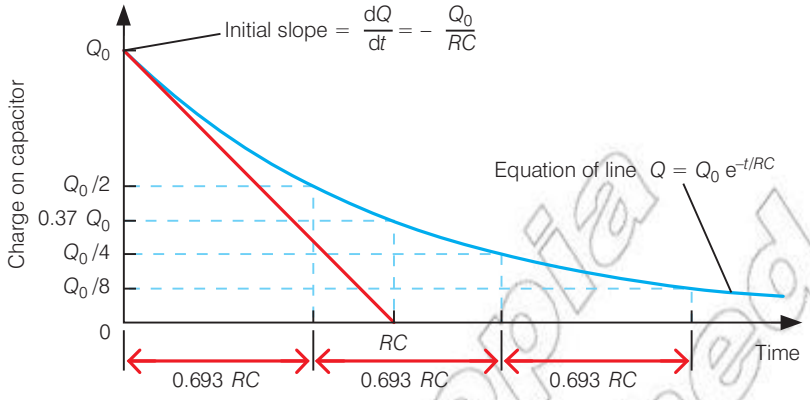


Figure 4.68 *The curve of the increase of charge on a capacitor has similar characteristics to the decay curve for the same capacitor. Notice that the increase in charge has a constant half-life of RC $\log_e 2$. The graph also illustrates how the capacitor would fully charge to a charge Q_f in time RC if it continued to charge at its initial rate, and how the actual charge on it after this time is $(1 - 0.37)Q_f = 0.63Q_f$.*

Uses of capacitors in a.c. circuits

Smoothing circuits

When a direct current (d.c.) is required from a mains a.c. circuit, a device called a rectifier is required, which will allow current through a load resistor in one direction only. A rectifier will supply a current that varies with time (Figure 4.69). In order to obtain a smooth d.c., a capacitor can be placed across the load resistor (Figure 4.70). During a cycle, the capacitor charges up when current is supplied and discharges through the load resistor when no current is supplied. The current with the capacitor in place is shown by the dotted line in Figure 4.69.

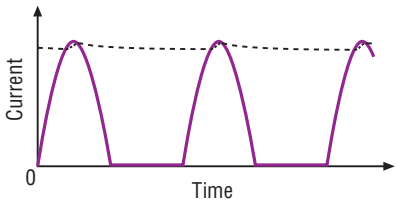


Figure 4.69 *A rectified alternating current (a.c.)*

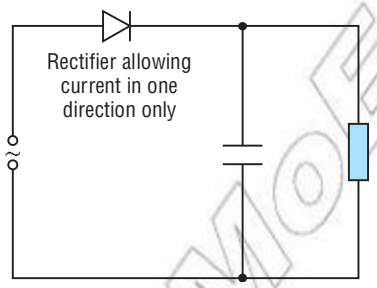


Figure 4.70 *A smoothing circuit*

Filter circuits

These are very important uses for capacitors. The effective resistance of a capacitor (called its reactance) varies with the frequency of the a.c. supply. This makes it possible to design circuits in which high-frequency signals travel in one direction in a circuit and low frequencies travel in another. For example, unwanted noise can be diverted from entering a loudspeaker. Another example of a filter circuit is shown in Figure 4.71. Here a supply is giving a fluctuating output. It is a combination of a.c. and d.c. The capacitor can charge up and discharge with the variations in output, but the d.c. component results in a fixed charge on the capacitor. The effect is to allow the a.c. through the capacitor but to block the d.c. Capacitors are frequently used to block d.c.

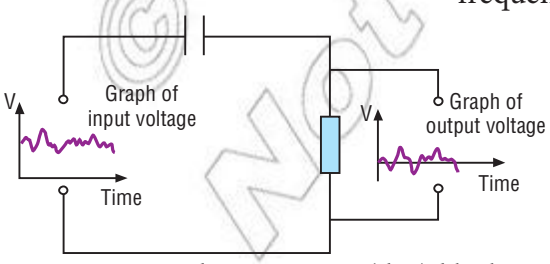
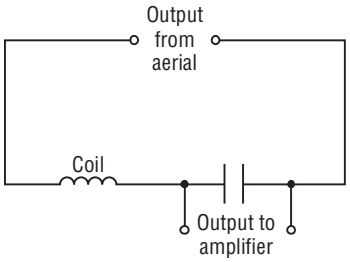


Figure 4.71 *A direct current (d.c.) blocking circuit*

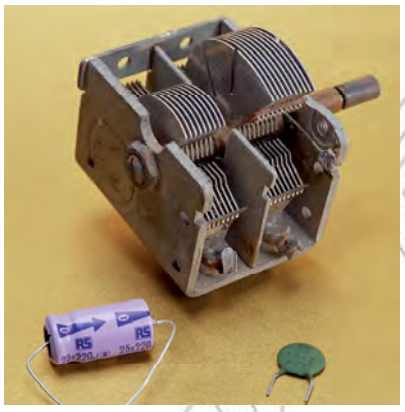
© MoE-Ethiopia Not to be republished

Tuning circuits

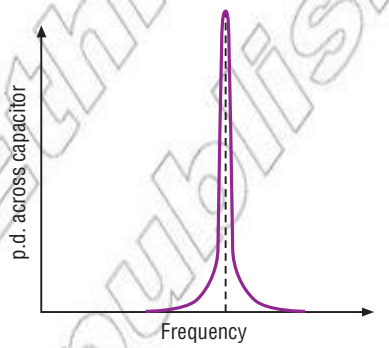
Combining a capacitor with a coil makes electrical resonance possible. The circuit is shown in Figure 4.72(a) in which an aerial is connected to a coil and a variable capacitor in series. In old radios the variable capacitor was like the one shown in Figure 72(b) and was right behind the tuning knob. The output across the capacitor varies with capacitance as shown in Figure 4.72(c). A signal of a particular frequency will give a large output. Notice that the shape of this graph is similar to that of one for mechanical resonance. This is electrical resonance, and as with mechanical resonance, the amplitude of the output can be much larger than the amplitude of the input.



(a)



(b)



(c)

Figure 4.72 *(a) A tuning circuit containing a coil and a capacitor; (b) a variable tuning capacitor; (c) the variation of output with capacitance*

Other uses

- Capacitor microphones: One plate of a capacitor in the microphone is free to vibrate. Its capacitance varies as sound waves cause it to vibrate. With a fixed p.d. across it, the charge on the capacitor must vary. The variations in the current to the microphone can be detected and amplified. A similar system is sometimes used for computer keyboards. When you depress a key you are not operating a switch but changing the capacitance of a capacitor.
- Displacement sensors: Moving your hand near a charged plate can alter the capacitance in a circuit and hence cause a small current. This can be used to sense the presence of a person.
- Preventing sparking: A capacitor across a switch will limit the damage caused by sparking and limit the amount of radio frequency interference a spark causes. It does this by allowing high-frequency a.c. to charge and discharge itself.
- As a counter in digital electronics (an integrator): If a digital signal, such as that shown in Figure 4.73, is used to charge up a capacitor, it will charge in steps and effectively count the number of pulses. It integrates. Other capacitor circuits can be used to differentiate.

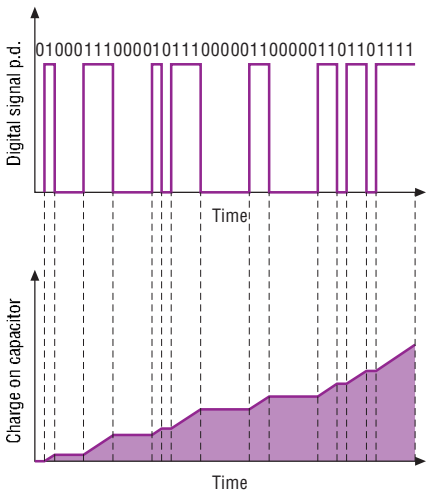


Figure 4.73 *A digital signal*

© Ministry of Education Ethiopia Not to be republished

Summary

In this section you have learnt that:

- The capacitance of a capacitor is given by $C = \frac{Q}{V}$ and is measured in farads (F).
- The capacitance of a parallel plate capacitor in a vacuum is given by $C = \frac{\varepsilon_0 A}{d}$.
- A dielectric material is an insulating material with properties that allow more charge to be stored. Dielectrics contain a number of dipoles.
- To determine the capacitance of a capacitor from its characteristics you can use $C = \frac{\varepsilon_0 \varepsilon_r A}{d}$.
- Connecting capacitors in series gives an effective capacitance equal to $\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$.
- Connecting capacitors in parallel gives an effective capacitance equal to $C_1 + C_2 + C_3 + \ldots$.
- The energy stored by a capacitor is equal to $\frac{1}{2}QV$.
- The electrical energy density of a parallel plate capacitor is given by $u = \frac{1}{2}\varepsilon_0 \varepsilon_r E^2$
- The discharge of a capacitor is given by $Q = Q_0 \exp(-t/RC)$, where RC is equal to the time constant of the discharging circuit.

Review questions

1. Define capacitance.
2. Use Gauss's law to show the capacitance of a parallel plate capacitor in a vacuum is given by $C = \frac{\varepsilon_0 A}{d}$.
3. Describe what happens to a dielectric when placed in an electric field and explain why they are used inside some capacitors
4. Figure 4.74 shows three capacitors connected to a d.c. supply.
   a) Calculate (i) the total capacitance of the system, (ii) the charge on and (iii) the p.d. across each capacitor.
   b) One of the 3.0 μF capacitors is replaced by one of an unknown value. The total capacitance of the system is 4.0 μF. Calculate the value of the unknown capacitor.
   c) For the capacitor system of part (b), calculate (i) the charge on and (ii) the p.d. across each capacitor.
5. A 1000 μF capacitor is charged from a 15.0 V d.c. supply through a two-way switch. The switch is thrown to connect it to an uncharged 500 μF capacitor as shown in Figure 4.75.
   a) Calculate (i) the initial and (ii) the final charge on the 1000 μF capacitor.

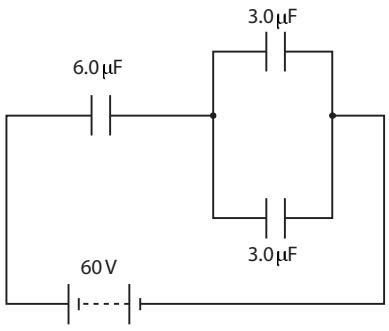


Figure 4.74

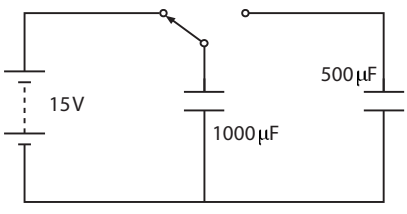


Figure 4.75

b) Calculate the change in p.d. across the 500 μF capacitor after the switch is thrown.

c) Calculate (i) the initial energy stored in the 1000 μF capacitor and (ii) the final energy stored in both capacitors.

6. A capacitor consists of two discs of metal 10 cm in diameter 1 mm apart in air. (Take $\varepsilon_0 = 8.85 \times 10^{-12}$ F/m.)

Calculate:

a) the capacitance of the capacitor

b) the charge on each plate of the capacitor if it is connected to a battery with an e.m.f. of 24 V.

7. Figure 4.76 shows a variable capacitor like that used in some radio tuners.

a) The capacitor can be thought of as several capacitors connected together. Are these connected in series or parallel?

b) What happens to the capacitance of the capacitor as the knob is turned anticlockwise?

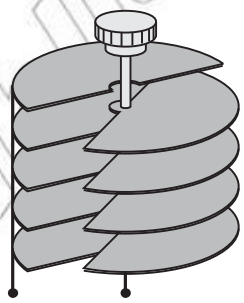


Figure 4.76

8. You have a sheet of polythene ($\varepsilon_r = 2.3$) 0.25 mm thick. If this polythene is to be used in a capacitor by sandwiching it between two sheets of aluminium foil, what area must the sheets have if the capacitor is to have a capacitance of

0.5 μF? (Take $\varepsilon_0 = 8.85 \times 10^{-12}$ F/m.)

9. A 10 μF capacitor is connected to a source of e.m.f. of $2 \times 10_4$ V.

a) Calculate the energy stored in the capacitor.

b) Comment on your answer.

10. A discharge lamp consists of gas at low pressure surrounding two electrodes. When it is not glowing, the gas in the lamp has a very high resistance. The gas does not conduct electricity until the potential difference across the lamp reaches a certain minimum value V_{min} (sometimes called the 'striking voltage'). Once conducting, the gas glows, and its resistance falls – it will continue to glow until the potential difference across it falls to around $0.75V_{min}$. Use this information to explain the observation in Figure 4.77, and sketch a graph of the reading on the voltmeter.

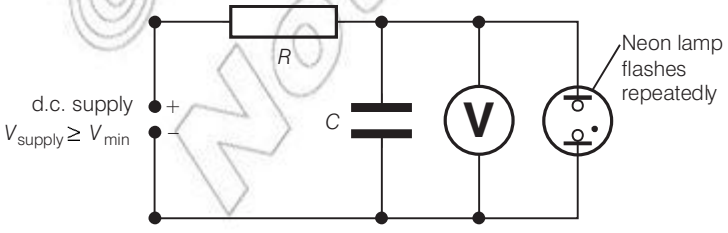


Figure 4.77

© MoE FDR Ethiopia Not to be republished

End of unit questions

1. Describe Millikan's experiment.

2. One practical arrangement for verifying Coulomb's law is to use a lightweight, negatively-charged, freely-suspended ball. It is repelled by the negative charge on a larger sphere that is held near it, on an insulated support. The small angle of deflection θ is then measured.

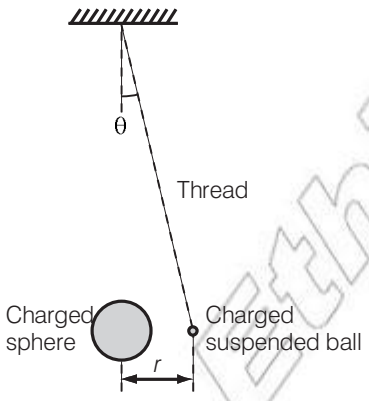


Figure 4.78

Draw a free-body force diagram for the suspended ball.

The weight of the ball is W. Show that the force of repulsion F on the suspended ball is given by

$F = W \tan\theta$

A student takes several sets of readings by moving the larger sphere towards the suspended ball in order to increase the mutual force of repulsion between them. He measures the angle of deflection θ and the separation distance r in each case. He then calculates the magnitude of the force F.

Here are some of his results.

Force $F/10^{-3}$ N			142	568
Distance $r/10^{-3}$ m	36.0	27.0	18.0	9.0

Calculate the values that you would expect the student to have obtained for the missing forces, assuming that Coulomb's law was obeyed.

Write your answers in a copy of the table.

Suggest why, in practice, it was necessary for the student to take measurements quickly using this arrangement.

3. Figure 4.79 shows a high-speed alpha particle entering the space between two charged plates in a vacuum.

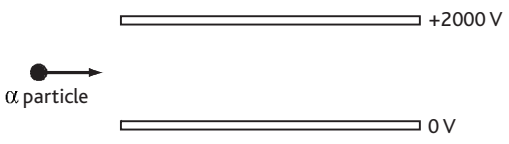


Figure 4.79

© MoE, FDR Ethiopia
Not to be republished

Add to a copy of the diagram the subsequent path of the alpha particle as it passes between the plates and well beyond them.

The gap between the plates is 10 mm. Calculate the magnitude of the electric force on the alpha particle as it passes between the plates.

4. Figure 4.80 shows two parallel plates with a potential difference of 3000 V applied across them. The plates are in a vacuum.

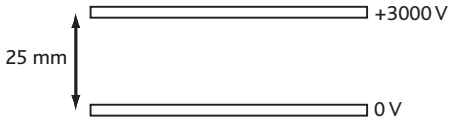


Figure 4.80

On a copy of the diagram sketch the electric field pattern in the region between the plates.

On the same diagram sketch and label two equipotential lines.

The plates are 25 mm apart. Show that the force experienced by an electron just above the bottom plate is about 2×10^{-14} N.

Copy and complete the graph to show how the force on the electron varies as the distance of the electron varies from the bottom plate to the top plate.



Figure 4.81

This force causes the electron to accelerate.

The electron is initially at rest in contact with the bottom plate when the potential difference is applied. Calculate its speed as it reaches the upper plate.

5. Define electric flux.

6. The spark plug of a car engine has two electrodes a distance 0.64 mm apart. The electric field between the electrodes must reach 3×10^6 Vm^{-1} if a spark is to be produced. What is the minimum potential required to do this?

7. A small sphere carrying a charge of 8 nC hangs between two metal plates a distance 10 cm apart. The mass of the sphere is 0.05 g. What potential difference between the plates will cause the sphere to make an angle of 10° with the vertical?

8. Refer to Figure 4.82. A charge of $+5 \times 10^{-6}$ C is placed at A, and one of $+1 \times 10^{-5}$ C at B. Calculate the magnitude and direction of the electric field at C.

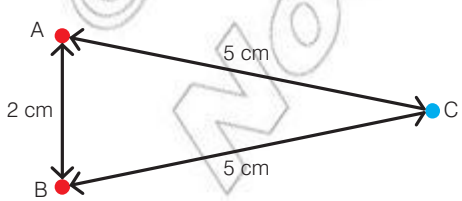


Figure 4.82

© Model FDRE Ethiopia Not to be republished

9. A raindrop of mass 50 mg and a charge of -1×10^{-10} C falls from a raincloud. The electric field between the cloud and the ground is 300 V/m.

   a) If the drop falls through a distance of 100 m, calculate the change in its:

      i) gravitational potential energy
      ii) electric potential energy.

   b) What electric field strength would be necessary to prevent the drop from falling? Is this likely to occur?

10 A pair of parallel flat metal plates are placed a distance of 10 mm apart. The plates are circular, with a radius of 10 cm. How much charge must be placed on each plate to produce an electric field of 500 V/m between them?

11. Define capacitance.

12. An uncharged capacitor of 200 μF is connected in series with a 470 kΩ resistor, a 1.50 V cell and a switch. Draw a circuit diagram of this arrangement.

    Calculate the maximum current that flows.

    Sketch a graph of voltage against charge for your capacitor as it charges. Indicate on the graph the energy stored when the capacitor is fully charged.

    Calculate the energy stored in the fully-charged capacitor.

12. Figure 4.83 shows a graph of charge against voltage for a capacitor.

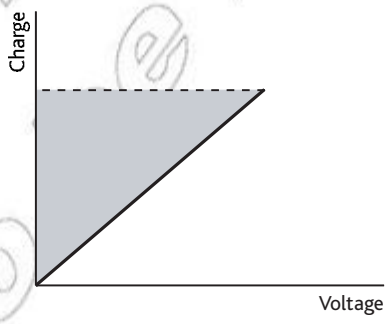


Figure 4.83

What quantity is represented by the slope of the graph?

What quantity is represented by the shaded area?

An electronic camera flash gun contains a capacitor of 100 μF which is charged to a voltage of 250 V. Show that the energy stored is 3.1 J.

The capacitor is charged by an electronic circuit that is powered by a 1.5 V cell. The current drawn from the cell is 0.20 A. Calculate the power from the cell and from this the minimum time for the cell to recharge the capacitor.

© MoE, FDR Ethiopia
Not to be republished

13. Calculate the capacitance between points A and B in Figure 4.84.

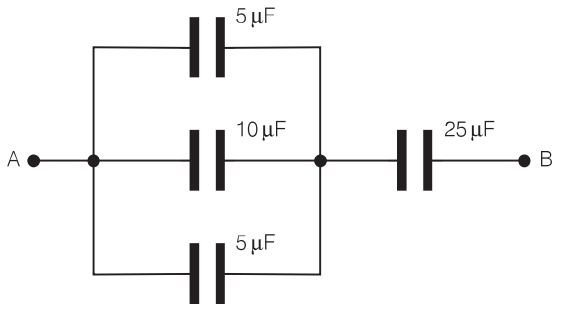


Figure 4.84

14. What capacitances can be made using four capacitors of 0.5 μF, 1.0 μF, 2.5 μF and 8.0 μF?

15. A 10 000 μF capacitor is charged by connecting it to a 12 V power supply. The capacitor is then discharged by connecting it to a length of copper wire of mass 1.5 g. If all the energy in the capacitor is dissipated as thermal energy in the wire, calculate the maximum temperature rise of the wire. (Take the specific heat capacity of copper as 390 J/kg K.)

16. A timing device is to be made using a capacitor that is to discharge through a fixed resistor. Time is to be measured using the potential difference across the capacitor. If a resistor of 5×10^4 W is available, what value capacitor will be appropriate for measuring times of around 10 s?

17. A charged capacitor will discharge slowly, even if its terminals are 'open circuit', that is, not connected to anything. This occurs because the dielectric between the two plates of the capacitor acts as a very poor conductor, allowing a very small leakage current to flow between the two plates, thus discharging them. A capacitor consists of two plates of area 50 cm^2, separated by a sheet of polythene 0.1 mm thick. The capacitor is briefly connected to a power supply, producing a potential difference between the plates of 15 V. Calculate:

    a) the capacitance of the capacitor

    b) the electrical resistance of the dielectric between the two plates

    c) the initial leakage current through the capacitor when it is disconnected from the power supply

    d) the time taken for the p.d. across the capacitor plates to fall to 7.5 V. (Take ε_r for polythene as 2.3, ε_0 as 8.85×10^{-12} F/m and the resistivity of polythene as 2×10^{11} Ω m.)

© MoE FDRE Ethiopia Not to be republished

Steady electric current and circuit properties

Unit 5

Contents

Section	Learning competencies
5.1 Basic principles (page 199)	• Define the terms resistance, resistivity, conductivity, current density, drift velocity. • Define the units coulomb, volt, ohm, watt, joule. • Identify that current density is a vector quantity. • Express drift velocity in terms of current density, number of charge carriers per unit volume and elementary charge. • Explain how the sources of e.m.f. produce a p.d. • Express the relationship between e.m.f., terminal p.d. and internal resistance. • Analyse, in quantitative terms, circuit problems involving potential difference, current and resistance. • Compute the p.d. across a resistor in a circuit.
5.2 Kirchoff's rules (page 214)	• State Kirchoff's junction rule. • Identify that Kirchoff's junction rule is a consequence of the law of conservation of charge. • State Kirchoff's loop rule. • Identify that the loop rule is a consequence of the conservation of energy. • Use Kirchoff's rules to solve related circuit problems. • Identify the sign conventions appropriately in applying Kirchoff's rules. • Solve problems involving network resistors.
5.3 Measuring instruments (page 220)	• Describe how a galvanometer can be modified to measure a wide range of currents and potential differences. • Describe how shunt resistors are used to measure a wide range of currents and p.d. • Calculate shunt and multiplier values for use with a meter to give different current and voltage ranges. • Solve problems in which a meter resistance is involved. • Identify and appropriately use equipment for measuring potential difference, electrical current and resistance (e.g. use multimeters and a galvanometer to make various measurements in an electrical circuit, use an oscilloscope to show the characteristics of the electrical current).
5.4 The Wheatstone bridge and the potentiometer (page 226)	• Explain the principle of the Wheatstone bridge and solve problems involving it. • Explain the principle of the potentiometer and how it can be used for measurement of e.m.f., p.d., resistance and current. • Solve problems involving potentiometer circuits.

5.1 Basic principles

By the end of this section you should be able to:

- Define the terms resistance, resistivity, conductivity, current density, drift velocity.
- Define the units coulomb, volt, ohm, watt, joule.
- Identify that current density is a vector quantity.
- Express drift velocity in terms of current density, number of charge carriers per unit volume and elementary charge.
- Explain how the sources of e.m.f. produce a p.d.
- Express the relation between e.m.f., terminal p.d. and internal resistance.
- Analyse, in quantitative terms, circuit problems involving potential difference, current and resistance.
- Compute the p.d. across a resistor in a circuit.

An electric current is a flow of charge. If you compare an electric current with water, a small current is like a trickle passing through a pipe; a really large current is like a river in flood.

The rate of flow of electric charge – that is, the electric current – is measured in amperes (A). The ampere is one of the fundamental units of the SI system. This means that the size of the ampere is not fixed in terms of other units: we simply compare currents to a 'standard ampere'.

An ampere is quite a sizeable flow of charge, especially in electronic circuits, so we also often deal in milliamperes (mA, thousandths of an ampere, 10^{-3} A) or even microamperes (μA, millionths of an ampere, 10^{-6} A).

If charge flows at a rate of one ampere, and continues to flow like that for a second, then the total amount of charge that has passed is one **coulomb**. A current of 3 A, for example, is a flow rate of 3 coulombs of charge every second (3 C/s). With that current, therefore, it should be obvious that in 10 seconds a total of 30 C of charge will pass, or that to supply 12 C of charge the current must flow for 4 seconds.

The formula linking amperes, coulombs and seconds is:

$Q = It$

where Q stands for the quantity of charge which passes when a current I flows for a time t. The units are

Q/coulombs (C)

I/amperes (A)

t/seconds (s)

DID YOU KNOW?

If two wires are placed 1 m apart and carry the same current, if the force between them is 2×10^{-7} N then the current flowing in each wire is 1 A.

KEY WORDS

coulomb *If charge flows at a rate of one ampere, and continues to flow like that for a second, then the total amount of charge that has passed is one coulomb.*

© Wolf Fm Namibia Not to be republished

Worked example 5.1

120 C of charge passes in 1 minute. What is the current?

Q (C)	I (A)	t (s)
120	?	60

Use $Q = It$

$$I = \frac{Q}{t} = \frac{120\ \text{C}}{60\ \text{s}} = 2\ \text{C/s} = 2\ \text{A}$$

Worked example 5.2

How long will a current of 5 A take to pass 100 C of charge?

Q (C)	I (A)	t (s)
100	5	?

Use $Q = It$

$$t = \frac{Q}{I} = \frac{100\ \text{C}}{5\ \text{A}} = 20\ \text{s}$$

Conduction electrons

We need to look in detail at the structure of an atom of a material that allows electric current to flow (a conducting material). All its positive charges are located in the central part, the nucleus. Each chemical element has a different number of positive charges in its nucleus, from one to nearly a hundred. Copper, for example, has 29 positive charges in the nucleus.

An uncharged copper atom must have 29 negative electrons as well. These electrons orbit round the nucleus, and each one has its own path. The first two electrons orbit in an innermost shell, the next eight fill a second shell and the following 18 complete the third shell. That leaves a solitary electron as the first member of a new fourth shell.

This electron in the outer shell allows copper to conduct electricity. It is comparatively easy to remove this electron, changing an uncharged copper atom into what we call a copper ion with an overall single positive charge.

In a copper wire the atoms are packed close together, as they are in any other solid. In each copper atom 28 of the electrons are still firmly bound in orbit around their nucleus, fixed in its place in the solid. The electrons in the outer shell remain trapped within the metal as a whole, but are free to drift about inside it. We call them the conduction electrons (see Figure 5.1). (You will learn more about the structure of the atom in section 8.2.)

© MoE Not to be republished

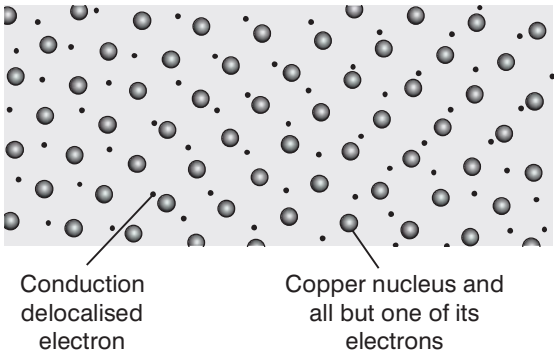


Figure 5.1 *Conduction electron in copper metal*

Conductivity, resistivity and resistance

Different materials have different numbers of conduction electrons. **Conductivity** is a way of measuring a material's ability to allow an electric current to flow. It is given the symbol σ and its units are Siemens per metre (S/m).

The inverse of conductivity is **resistivity**. Resistivity is a measure of how much a material resists the flow of an electric current. It is given the symbol ρ and its units are ohm metres (Ω m). A material with a high conductivity will have a low resistivity and a material with a high resistivity will have a low conductivity.

Resistance is a property of a material that controls the amount of current that flows through it. It is measured in **ohms** (Ω).

The resistance of a metal wire at a given temperature is determined by three factors:

- Its length l, in metres – the resistance is proportional to l, so if the length doubles so does the resistance.
- Its area of cross-section A, in m^2 – the resistance is inversely proportional to A, so a wire with twice the cross-sectional area will have only half the resistance.
- The resistivity (in ohm metres) of the material from which the wire is made. A material with higher resistivity will have a higher resistance.

Resistivity and resistance are thus related by the equation

$$R = \frac{\rho l}{A}$$

KEY WORDS

conductivity *a way of measuring a material's ability to allow an electric current to flow*
resistivity *a measure of how much a material resists the flow of an electric current*
resistance a *property of a material that controls the amount of current that flows through it*
ohm *the unit of resistance*

DID YOU KNOW?

The resistivities of most metals are in the range 10^{-7} to 10^{-8} Ω m. The ones with the larger resistivities conduct electricity less well. For an insulator such as dry polythene the resistivity may be as high as 10^{15} Ω m. Those are two extremes, conductor and insulator. There are just a few materials in between: germanium at room temperature, for instance, may display a resistivity of around 0.001 Ω m.

© MoE-FDRE Ethiopia Not to be republished

Worked example 5.3

What will the resistance of a copper cable be if it has a cross–sectional area of 1 cm^2 and a length of 2 km? The resistivity of copper is $2 \times 10^{-8}\ \Omega$ m.

Be careful over the units.

$l = 2\ km = 2 \times 10^3\ m$

$A = 1\ cm^2 = 1 \times 10^{-4}\ m^2$ (since there are 100×100 square centimetre in a square metre)

$R\ (\Omega)$	$\rho\ (\Omega\ m)$	$l\ (m)$	$A\ (m^2)$
?	2×10^{-8}	2×10^3	1×10^{-4}

Use

$$R = \frac{\rho l}{A}$$

Putting in the values, we get

$$R = \frac{2 \times 10^{-8} \times 2 \times 10^3}{1 \times 10^{-4}} = 0.4\ \Omega$$

Worked example 5.4

Constantan has a resistivity of $47 \times 10^{-8}\ \Omega$ m. How much of this wire is needed to make a 10 Ω resistor if the diameter is 0.5 mm?

Be careful with the units.

Work out the radius in metres:

$r = 0.25 \times 10^{-3}$ m

Now work out the area:

$A = \pi r^2 = \pi \times (0.25 \times 10^{-3})^2 = \pi \times 6.25 \times 10^{-8}\ m^2 = 1.96 \times 10^{-7}\ m^2$

$R\ (\Omega)$	$\rho\ (\Omega\ m)$	$l\ (m)$	$A\ (m^2)$
10	47×10^{-8}	?	$1.96 \times 10^{-7}\ m^2$

Use

$$R = \frac{\rho l}{A}$$

This is rearranged to

$$l = \frac{RA}{\rho}$$

Put in the values

$$l = \frac{10 \times 1.96 \times 10^{-7}}{47 \times 10^{-8}} = 4.17\ m$$

© MoE Not for sale

Drift velocity

Even when no current flows through a piece of copper, the free electrons are moving rapidly about. Their speed is about 10^6 m/s, or 3000 times the speed of sound in air. However, since they are moving at random, there is no *net* flow of electrons in any particular direction and so there is no current (see Figure 5.2).

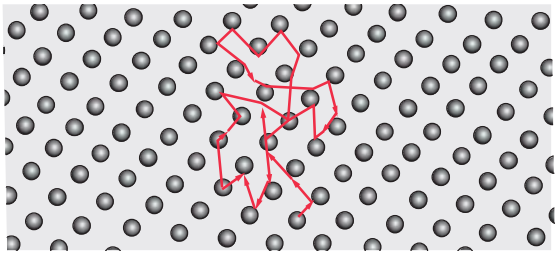


Figure 5.2 *Path of conduction electron when there is no current: no general drift of electrons*

When an electric field in the form of a voltage is applied, the electrons gain an additional velocity, so that there is a net flow along the wire (see Figure 5.3). This extra velocity is called their **drift velocity**.

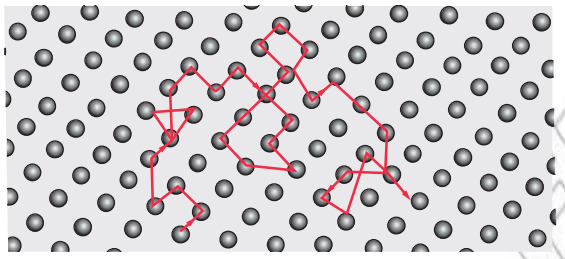


Figure 5.3 *Path of conduction electron when there is a current: general drift of electrons*

Activity 5.1: An analogy for drift velocity

Your teacher will show you an analogy for drift velocity. You will have a maze of nails on a board and some ball bearings, of diameter of 0.6 cm or less. The ball bearings moving through the maze of nails are an analogy for the movement of charge carriers in the wire. You can simulate a change in voltage by changing the angle of the board. Try varying the number of ball bearings (charge carriers) and the angle of the board (voltage). What happens in each case?

Current density

Current density is a vector quantity, which means it has both magnitude and direction. Its magnitude is the current per cross-sectional area. Its units are A/m^2 and it is given the symbol J.

There is a common approximation to the current density which assumes that the current is proportional to the electric field, E, that produces it. The relationship is

$J = \sigma E$

KEY WORDS

drift velocity *the average velocity that an electron reaches when an electric field is applied across a conductor*
current density *is a vector quantity, which means it has both magnitude and direction. Its magnitude is the current per cross-sectional area.*

DID YOU KNOW?

Current density is important to the design of electrical and electronic systems.

Circuit performance depends strongly upon the designed current level, and the current density then is determined by the dimensions of the conducting elements. For example, as integrated circuits are reduced in size, despite the lower current demanded by smaller devices, there is trend toward higher current densities to achieve higher device numbers in ever smaller chip areas.

© Ethio-Korea Publishing House (EKPH) DR not to be republished

where J is the current density, σ is the electrical conductivity of the material and E is the electric field.

Worked example 5.5

Find the approximate current density when an electric field of 5 V/m is applied to a copper conductor. The conductivity of copper is 59.6×10^6 S/m.

J (A/m^2)	σ (S/m)	E (V/m)
?	59.6×10^6	5

Use $J = \sigma E$

$= 59.6 \times 10^6 \times 5$

$= 2.98 \times 10^8$ A/m^2

Drift velocity and its relationship to current density

We can find an equation for drift velocity by beginning with the definition of current:

$$\frac{\Delta Q}{\Delta t}$$

where ΔQ is the small amount of charge that passes through an area in a small unit of time, Δt.

However,

ΔQ = (number of charged particles) × (charge per particle)

$= (nA\Delta x)q$

where

n is the number of charge carriers per unit volume

A is the cross-sectional area

Δx is a small length along the wire

q is the charge of the charge carriers.

We know that under the influence of an electric field in the wire, the charge carriers gain an average velocity in a specific direction, the drift velocity, v_d. Since $\Delta x = v_d \Delta t$, the above equation becomes

$\Delta Q = (nAv_d\Delta t)q$

We now put this back into the original equation for current and rearrange it so that drift velocity is the subject:

$$v_d = \frac{I}{nqA}$$

However, current density J is current per cross-sectional area, so

$$J = \frac{I}{A}$$

If we substitute this into the equation for v_d, we find that

$$v_d = \frac{J}{nq}$$

© MoE. FDR Not to be republished

where v_d is the drift velocity, J is the current density, n is the number of charge carriers per unit volume and q is the elementary charge on the charge carriers.

Worked example 5.6

Find the number of charge carriers per unit volume in a copper wire of cross-sectional area 1 mm^2 carrying a current of 3 A if the drift velocity of the charge carriers is 0.00028 m/s. (The elementary charge is 1.6×10^{-19} C.)

J (A/m^2)	n	q (C)	v_d (m/s)
$3/1 \times 10^{-6} = 3 \times 10^6$	?	1.6×10^{-19}	0.00028

Use $v_d = \frac{J}{nq}$

$$n = \frac{J}{qv_d}$$

$$= \frac{3 \times 10^6}{0.00028 \times 1.6 \times 10^{-19}} = 7 \times 10^{28}$$

Activity 5.2: Summarising your learning

To summarise what you have learnt in this unit so far, work in a small group to produce a poster showing how resistance, resistivity, conductivity, drift velocity and current density are related.

How does a source of e.m.f. produce a p.d.?

We know that when a voltage is connected across a piece of copper, it pushes the free electrons so that they flow through the metal and produce an electric current. The electrons start to flow instantaneously because the free electrons are already spread through the wire.

As soon as the voltage is applied, there is an electromotive force on all the electrons, which gets them moving. It's a bit like a bicycle chain. As soon as you start pedalling, the back wheel starts to turn. The force on the back wheel is instantaneous even though the individual links are travelling at a visible speed. However, because the links are already spread around the chain 'circuit' they all start to move at the same time.

Electrical circuits transfer energy from batteries to the other components. The chemicals in the battery are a store of energy. When the circuit is complete the energy from the battery pushes the current around the circuit and transfers the energy to the component, which can then work. The energy or push that the battery gives to the circuit is called is called the voltage or electromotive force (e.m.f.) of the battery. It is measured in **volts**, which have the symbol V.

The higher the voltage is the greater the amount of energy that can be transferred. Voltage is a measure of the difference in electrical energy between two parts of a circuit. Because the energy is transferred by the component there must be more energy entering the component than there is leaving the component. Voltage is sometimes called potential difference (p.d.). Potential difference measures the difference in the amount of energy the current is

KEY WORDS

volt *a measurement of voltage or electromotive force, defined as joules per coulomb*

© Not to be republished

KEY WORDS

joule *One joule is the energy exerted by a force of one newton acting to move an object through a distance of one metre*

carrying either side of the component. This voltage drop across the component tells us how much energy the component is transferring.

Potential difference is defined as energy per unit charge. The unit of potential difference is the volt (V). Using the definition, we can define the volt as **joules** per coulomb.

1 V = 1 J/C

Activity 5.3: Remembering Ohm's law and power

In a small group, see what you can remember about Ohm's law from Grade 10. (Hint: it connects p.d., current and resistance.)

Now consider a current I flowing for t seconds in a component. The charge that flowed led to E joules being dissipated in the component. Use this information to derive expressions for the power dissipated in a resistor.

Activity 5.4: Saving on your electricity bill

Fluorescent bulbs deliver the same amount of light as conventional bulbs but use much less power. If 1 kW hour costs 10c, estimate the amount of money that would be saved in your household every month if you replaced all the 75 W incandescent bulbs by 15 W fluorescent ones.

Activity 5.5: Plotting V–I characteristics for an unknown resistance

Set up the circuit shown in Figure 5.4.

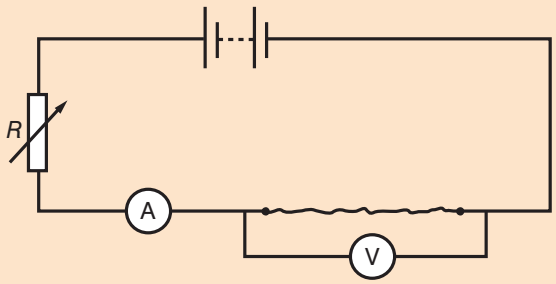


Figure 5.4 *Circuit to determine an unknown resistance*

Use 50 cm constantan wire as the unknown resistance. Vary the current through the unknown resistance by altering the value of the variable resistor, R. Record values of the p.d. across the unknown resistor and the current through it. Plot a graph of p.d. against current. Use the graph to calculate the resistance per cm of the wire. (The resistivity of constantan is 4.9×10^{-7} Ω m.)

The relationship between e.m.f., terminal p.d. and internal resistance

Suppose you short-circuit a battery. This means that you join its two terminals by a circuit that effectively has no resistance; a short piece of very thick copper wire, for instance. The battery has an e.m.f. V, but the circuit apparently has no resistance R. What happens then? Does the current increase without limit?

The thing we are forgetting is that the battery has to pump the charge round the whole circuit, and that includes the bit within the battery as well as the outside circuit. The internal resistance varies

© MoE Not to be republished

a lot between the different sorts of batteries, but that is what finally sets a limit to the current they can supply.

A 1.5 V torch battery typically has an internal resistance of up to an ohm. This means that even if you short-circuit the battery there is still that ohm of resistance present. The biggest current it can deliver is given by $I = \frac{V}{R} = \frac{1.5}{1} = 1.5$ A.

There is a formula that relates e.m.f. (E), terminal p.d. (V) and the internal resistance of the source (r):

$E = V + Ir$

Think about this...

What is the ratio of the internal resistance to the total resistance? How is this related to the answer to part c) of worked example 5.7? Is this result always true for such circuits? Try some circuits of your own and see!

Worked example 5.7

The diagram shows a series circuit with an internal resistance, r.

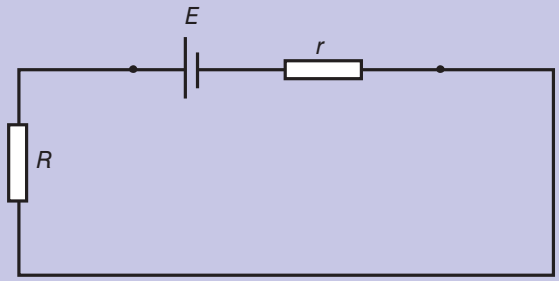


Figure 5.5 *A circuit with internal resistance*

The battery has an e.m.f. of 12 V and an internal resistance of 3 Ω. Calculate:

a) the current it supplies to the resistor, R, with value 12 Ω

b) the power used in the external resistor

c) the percentage of the total power wasted in the internal resistance.

a)

Total resistance $R + r$ (Ω)	E.m.f. (V)	Current I (A)
15	12	?

Use Ohm's law $V = IR$

In this case

$$\text{Current} = \frac{\text{e.m.f.}}{\text{total resistance}}$$

$$I = \frac{12}{15} = 0.8 \text{ A}$$

b)

Power (W)	Current (A)	R (Ω)
?	0.8	12

Use power $= I^2R$

$= 0.8^2 \times 12$

$= 7.68$ W

Activity 5.6: $E = V + Ir$

Set up a circuit using a battery and an external resistor like the one shown in worked example 5.7. Take measurements to verify the relationship $E = V + Ir$ and the result found in part c) of the worked example.

c) We need to find the power in internal resistance, then total power, then percentage of total power wasted in internal resistance

For internal resistance:

Power (W)	Current (A)	r (Ω)
?	0.8	3

Use power $= I^2R$

$= 0.8^2 \times 3$

$= 1.92$ W

Total power $= 1.92 + 7.68$ W

$= 9.6$ W

Percentage wasted in internal resistance $= \frac{1.92 \times 100}{9.6} = 20\%$

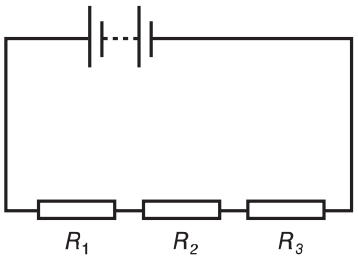


Figure 5.6 *Resistors in series*

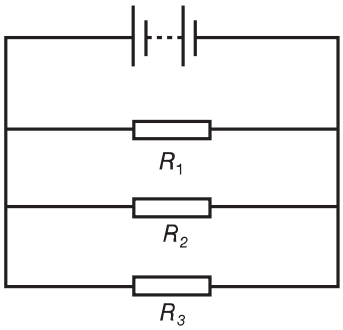


Figure 5.7 *Resistors in parallel*

Combining resistors

In Grade 10, you learnt that resistors can be combined in two ways: in series and in parallel. You used the following formulae.

For resistors in series (as shown in Figure 5.6), the total resistance R_T is given by:

$$R_T = R_1 + R_2 + R_3$$

For resistors in parallel (as shown in Figure 5.7), the total resistance R_T is given by:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Activity 5.7: Verification of the laws of combinations of resistors

Set up circuits to verify the laws for combination of resistors in series and in parallel.

Analysing circuits

You can use Ohm's law and the results for combinations of resistors to analyze circuit problems involving potential difference, current and resistance. You will learn more about how measuring instruments work in section 5.3.

© MoE: FDRE Not to be republished

Worked example 5.8

Figure 5.8 shows part of an electronic circuit. Calculate i) the p.d. between points A and B, and ii) the current in the 2.2 kΩ resistor when the switch S is:

a) open

b) closed.

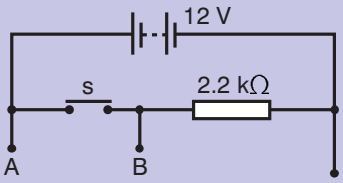


Figure 5.8 *Part of an electronic circuit*

When the switch is open the circuit is not complete. The entire p.d. will be between A and B and there will be no current in the resistor.

So the answers are i) 12 V ii) 0 A

When the switch is closed the circuit is complete. There is very little resistance between points A and B so the entire p.d. will be across the resistor.

For the resistor:

p.d. (V)	R (Ω)	I (A)
12	2.2×10^3	?

Use Ohm's law

$$I = \frac{V}{R}$$

$$= \frac{12}{2.2 \times 10^3}$$

$$= 5.5 \text{ mA}$$

Worked example 5.9

When an ammeter is added to a circuit to measure the current it acts as a series resistor of resistance R. The circuit in Figure 5.9 consists of a 12 V supply of negligible internal resistance connected to two equal resistors and the ammeter A.

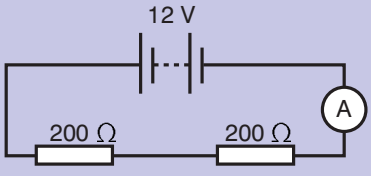


Figure 5.9 *Ammeter of unknown resistance*

a) Write down the total resistance in the circuit i) before ii) after the ammeter is added.

b) Show that the current before the ammeter is added is 30 mA.

c) The ammeter reads 24 mA when it is in the circuit. Calculate its resistance R.

Ethiopia published

a) i) Use $R_T = R_1 + R_2$

$= 200 + 200 = 400\ \Omega$

ii) Use $R_T = R_1 + R_2 + R_3$

$= (200 + 200 + R)\ \Omega$

b)

p.d. (V)	R (Ω)	I (A)
12	400	?

Use Ohm's law

$$I = \frac{V}{R}$$

$$= \frac{12}{400}$$

$= 0.03$ A $= 30$ mA

c)

p.d. (V)	R (Ω)	I (A)
12	$400 + R$	0.024

Use Ohm's law

$$400 + R = \frac{12}{0.024}$$

$400 + R = 500$

$R = 100\ \Omega$

Worked example 5.10

Two resistors of resistance 20 Ω and 40 Ω are connected in series to a 6.0 V cell. Calculate:

a) the total resistance in the circuit

b) the current in the circuit

c) the p.d. across the 40 Ω resistor.

a) Use

$R_T = R_1 + R_2$

$= 20 + 40$

$= 60\ \Omega$

b)

p.d. (V)	R (Ω)	I (A)
6.0	60	?

Use Ohm's law

$$I = \frac{V}{R}$$

$$= \frac{6}{60}$$

$= 0.1$ A

© MoE Not to

c)

p.d. (V)	R (Ω)	I (A)
?	40	0.1

Use Ohm's law $V = IR$

$= 0.1 \times 40$

$= 4$ V

Think about this...

Why is the value of two equal resistors in parallel half the value of each of them? Prove this mathematically.

Worked example 5.11

Calculate the total resistance of the network of resistors shown in Figure 5.10.

Deal with the two parallel sections first.

For the left hand pair use

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$\frac{1}{R_T} = \frac{1}{100} + \frac{1}{25}$$

$$= \frac{1}{100} + \frac{4}{100} = \frac{5}{100}$$

$$R_T = \frac{100}{5} = 20\ \Omega$$

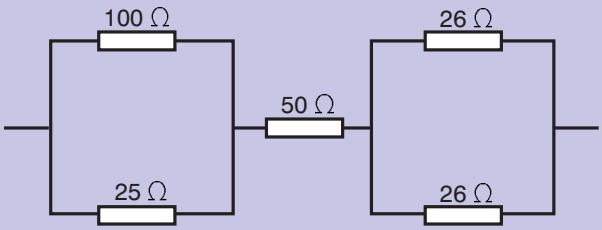


Figure 5.10 *resistor network*

The right hand pair are equal so their total resistance is half the value of each, i.e. $R_T = 13\ \Omega$.

The total resistance of the network is therefore

$20 + 50 + 13 = 83\ \Omega$

Summary

In this section you have learnt that:

- Conductivity is a way of measuring a material's ability to allow an electric current to flow.
- Resistivity is a measure of how much a material resists the flow of an electric current. It is the inverse of conductivity.
- Resistance is a property of a material that controls the amount of current that flows through it.
- Current density is a vector quantity. This means it has both magnitude and direction. Its magnitude is the current per cross-sectional area. It is given the symbol J. An approximation is $J = \sigma E$ where σ is the conductivity of the material and E is the electric field applied.
- Drift velocity is the average velocity that an electron reaches when an electric field is applied across a conductor.

- Drift velocity, *vd*, can be expressed in terms of current density, J, number of charge carriers per unit volume, n, and elementary charge, q, using the equation $vd = \frac{J}{nq}$.
- If charge flows at a rate of one ampere, and continues to flow like that for a second, then the total amount of charge that has passed is one coulomb.
- One volt is one joule per coulomb.
- One ohm is the resistance of a material when a p.d. of 1 volt is applied across the ends of the material and a current of 1 A flows.
- One joule is the energy exerted by a force of one newton acting to move an object through a distance of one metre.
- One watt is defined as 1 joule per second.
- Sources of e.m.f. produce a potential difference because they produce a difference in electrical potential at either end of a conductor.
- E.m.f. (E), terminal p.d. (V), and internal resistance (r) are related by the equation $E = V + Ir$.
- Ohm's law can be summarised using the equation $V = IR$. This can be used to analyse circuits and solve circuit problems involving potential difference, current and resistance.
- The power dissipated in an electrical component is $P = IV = I^2R$.

Review questions

1. When a small torch is switched on, the current drawn from the cell is 0.2 A.
   a) Calculate the charge passing a point in the bulb filament when the torch is switched on for 10 minutes.
   b) How many electrons drift past the point in this time? (The charge on one electron is 1.6×10^{-19} C.)
2. The current in a lightning strike is 7500 A. The strike lasts for 240 ms. Calculate
   a) the charge, in C, which flows in the strike to the ground
   b) the number of electrons transferred to the ground.
3. a) The resistivity of zinc is 5.9×10^{-8} Ω m. Copper is a better conductor than zinc. Does this mean that copper has a higher or lower resistivity than zinc?
   b) A 1 m length of copper wire of diameter 0.4 mm has a measured resistance of 0.13 Ω. What value does this give for the resistivity of copper?

© MoE Not to be

c) Why might this value be different from the actual value of the resistivity of copper?

4. Find the approximate current density when an electric field of 12 V/m is applied to a silver conductor. The conductivity of silver is 63.0×10^6 S/m.

5. Find the number of charge carriers per unit volume in a copper wire of cross-sectional area 2 mm^2 carrying a current of 1.5 A if the drift velocity of the charge carriers is 0.00028 m/s. (The elementary charge is 1.6×10^{-19} C.)

6. Figure 5.11 shows a series circuit with an internal resistance, *r*.

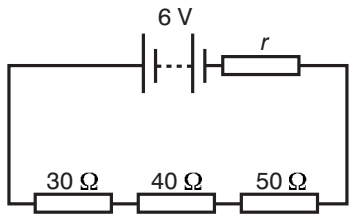


Figure 5.11

The battery has an e.m.f. of 6 V and an internal resistance, *r*, of 1.5 Ω. Calculate:

a) the current it supplies to the external resistors

b) the power used in the external resistors

c) the percentage of the total power wasted in the internal resistance

d) the p.d. across the 40 Ω resistor.

7. Figure 5.12 shows a 12 V battery of negligible internal resistance connected to three resistors and an ammeter.

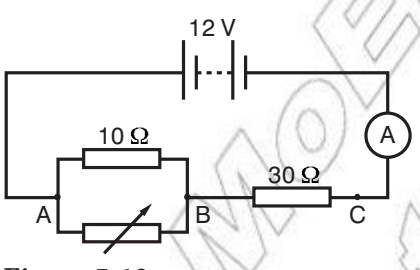


Figure 5.12

a) The variable resistor is set to it maximum value of 15 Ω. Calculate the resistance between points i) A and B; ii) A and C.

b) Calculate the maximum and minimum readings on the ammeter when the variable resistor is set to 0 Ω and 15 Ω.

© MoE FDR Ethiopia Not to be republished

DID YOU KNOW?

Gustav Robert Kirchhoff (1824–1887) was a German physicist who contributed to the fundamental understanding of electrical circuits, spectroscopy and the emission of black-body radiation by heated objects. He coined the term "black body" radiation in 1862, and two sets of independent concepts in both circuit theory and thermal emission are named "Kirchhoff's laws" after him. Kirchhoff formulated his circuit laws in 1845 while he was still a student. He completed this study as a seminar exercise; it later became his doctoral dissertation.

5.2 Kirchoff's rules

By the end of this section you should be able to:

- State Kirchoff's junction rule.
- Identify that Kirchoff's junction rule is a consequence of the law of conservation of charge.
- State Kirchoff's loop rule.
- Identify that the loop rule is a consequence of the conservation of energy.
- Use Kirchoff's rules to solve related circuit problems.
- Identify the sign conventions appropriately in applying Kirchoff's rules.
- Solve problems involving network resistors.

Kirchoff's junction rule

When an electric current arrives at a junction, the current divides into two or more parts, with some electrons going in one direction and the rest going along the other paths. This is true no matter how complicated the junction or the circuit may be. Electrons cannot appear or disappear so charge is said to be conserved.

A battery does not produce electric charge, it simply pumps the charge around the circuit. A large number of electrons enter the battery at the positive terminal every second, and the same number leave the battery at the negative terminal every second. Similarly, the rate at which electrons arrive at one end of a wire is exactly the same as the rate at which they leave the other. This is all summarised in Kirchoff's junction rule which states that

the total current flowing into a point is equal to the total current flowing out of that point.

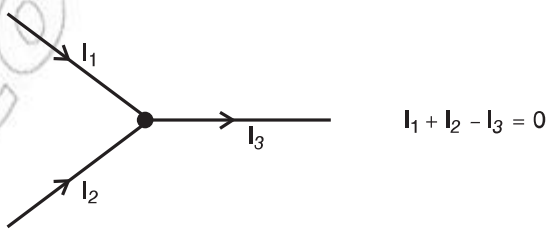


***Figure 5.13** Kirchoff's junction rule*

In Figure 5.13, we can see that $I_3 = I_1 + I_2$

This can be written as $I_1 + I_2 - I_3 = 0$.

Notice that I_3 has a negative sign. By convention, currents going into a junction are positive but currents leaving a junction are negative. The sum of the currents at any junction is zero.

© MoE. FREE Not to be republished

Worked example 5.12

Figure 5.14 shows part of a circuit network. State the value of the current in each of the resistors A, B and C

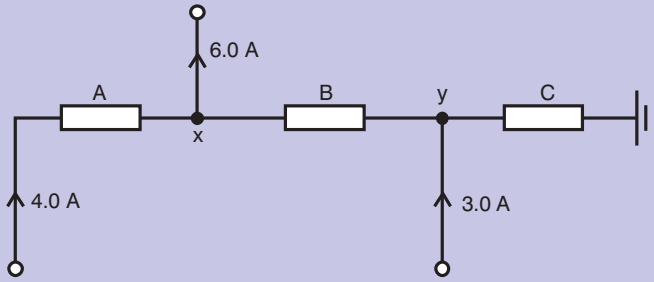


Figure 5.14

Resistor A has the full 4.0 A flowing through it.

For resistor B, we need to consider junction x. There are 6.0 A leaving X, which means a total of 6.0 A must enter x. Since 4.0 A enter from the left, Kirchoff's junction rule states that 2.0 A must enter from the right, as shown in Figure 5.15.

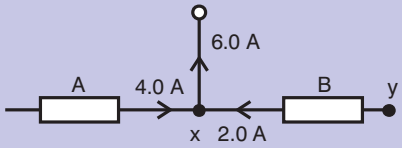


Figure 5.15

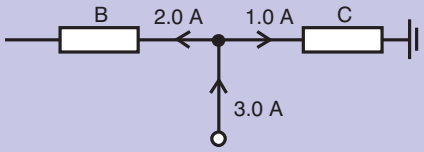


Figure 5.16

For resistor C, we need to consider junction y. There are 3.0 A flowing into the junction, so a total of 3.0 A must leave the junction. From above, we know that 2.0 A flow through B, leaving 1.0 A to flow through C, as shown in Figure 5.16.

Kirchoff's loop rule

We can consider e.m.f. to be energy per unit charge transferred into electrical energy and p.d. to be energy transferred from electrical energy. We know that energy is always conserved. In a circuit, the electrical energy supplied by the battery is used in the circuit – no surplus energy arrives back at the battery. Kirchoff's loop rule recognises this and is stated as

in any closed loop in a circuit the sum of the e.m.f.s is equal to the sum of the p.d.s

We can use this rule to derive the equation for the sum of resistors in parallel that we met in section 5.1. Consider the circuit shown in Figure 5.17.

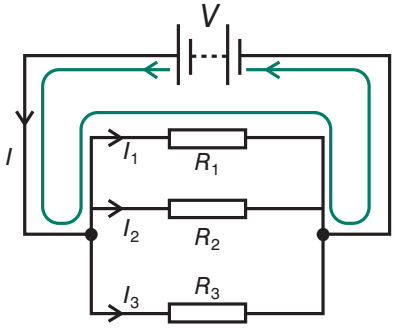


Figure 5.17

We assume that the battery has negligible internal resistance. If we apply Kirchoff's loop rule to the complete loop from the battery to R_1 and back again to the battery, as shown in the diagram, then the p.d. across the resistor equals the e.m.f. of the battery. This is true for each resistor, so if we now apply Kirchoff's junction rule we get

$$I = I_1 + I_2 + I_3 = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

Since $I = \frac{V}{R}$ where R is the total resistance, we get

$$\frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

Now divide throughout by V and we get

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Worked example 5.13

Find the current that flows in each of the resistors in the circuit shown in Figure 5.18.

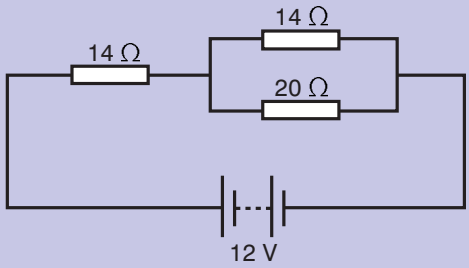


Figure 5.18

Start by drawing the diagram with the different currents marked in Figure 5.19 as shown.

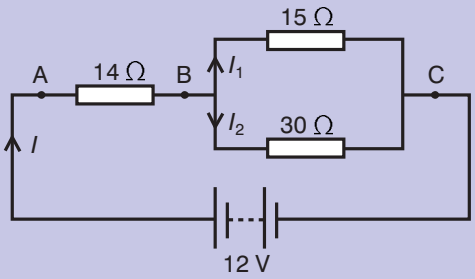


Figure 5.19

By applying Kirchoff's loop rule we can say that the p.d. between points A and C is 12 V.

Now we need to find the current I.

We can do this by first working out the effective resistance in the circuit. The resistance of the parallel combination is given by

$$\frac{1}{R} = \frac{1}{15} + \frac{1}{30} = \frac{2}{30} + \frac{1}{30} = \frac{3}{30}$$

$R = 10\ \Omega$

© MoE-Ethiopia Not to be published

So the total resistance = 10 + 14 = 24 Ω

So the current I is given by = $\frac{V}{R} = \frac{12}{24} = 0.5$ A

Now we need to find the p.d. between points B and C in order to find I_1 and I_2.

First we need to know the p.d. between A and B.

$V = IR = 0.5 \times 14 = 7$ V

Applying Kirchoff's loop rule again, this means that the p.d. between B and C must be 12 – 7 V = 5 V

So since $I = \frac{V}{R}$ we know that

$I_1 = \frac{5}{15}$ A and $I_2 = \frac{5}{30}$ A

We can now check that $I = I_1 + I_2$

$\frac{15}{30} = \frac{10}{30} + \frac{5}{30}$

This confirms our answer.

Worked example 5.14

Find the currents flowing in each of the resistors in this circuit.

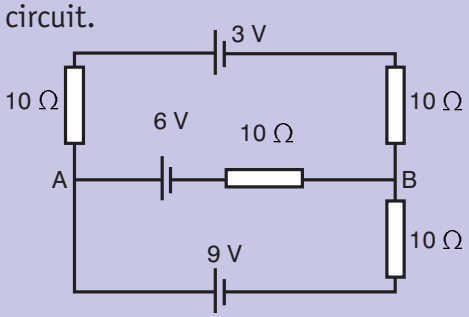


Figure 5.20a

First label the diagram as shown below.

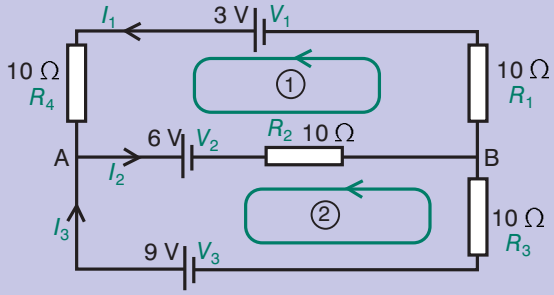


Figure 5.20b

Ethiopia Not to be republished

At junction A, from Kirchoff's junction rule, $I_1 + I_3 = I_2$ (1)

In loop 1, from Kirchoff's loop rule, $V_1 - I_1R_4 - V_2 - I_2R_2 - I_1R_1 = 0$

Substituting in the values we get $3 - 10I_1 - 6 - 10I_2 - 10I_1 = 0$

$-3 = 20I_1 + 10I_2$ (2)

In loop 2, $V_3 - V_2 - I_2R_2 - I_2R_3 = 0$

Substituting in the values we get

$9 - 6 - 10I_2 - 10I_2 = 0$

$3 = 20I_2$ (3)

From (3) $I_2 = \frac{3}{20}$ A = 0.15 A

Substitute this value into (2)

$-3 = 20I_1 + 10 \times 0.15$

$-3 = 20I_1 + 1.5$

$-4.5 = 20I_1$ (4)

From (4), $I_1 = -4.5/20$ A = –0.225 A

Note that the minus sign indicates that the current flows in the opposite direction to that shown on the diagram.

Substitute the values we have found into (1), $I_1 + I_3 = I_2$

$-0.225 + 0.15 = I_3$

-0.075 A $= I_3$

Note that the minus sign indicates that the current flows in the opposite direction to that shown on the diagram.

So the currents are $I_1 = -0.225$ A, $I_2 = 0.15$ A, $I_3 = -0.075$ A.

Summary

In this section you have learnt that:

- When an electric current arrives at a junction, the current divides into two or more parts, with some electrons going in one direction and the rest going along the other paths. This is true no matter how complicated the junction or the circuit may be. Electrons cannot appear or disappear so charge is said to be conserved.
- A battery does not produce electric charge, it simply pumps the charge around the circuit. A large number of electrons enter the battery at the positive terminal every second, and the same number leave the battery at the negative terminal every second. Similarly, the rate at which electrons arrive at one end of a wire is exactly the same as the rate at which they leave the other. This is all summarised in Kirchoff's junction rule.
- Kirchoff's junction rule states that the total current flowing into a point is equal to the total current flowing out of that point.

© MoE Not for sale

- By convention, currents going into a junction are positive but currents leaving a junction are negative. The sum of the currents at any junction is zero.
- We can consider e.m.f. to be energy per unit charge transferred into electrical energy and p.d. to be energy transferred from electrical energy. We know that energy is always conserved. In a circuit, the electrical energy supplied by the battery is used in the circuit – no surplus energy arrives back at the battery.
- Kirchoff's loop rule recognises this and is stated as in any closed loop in a circuit the sum of the e.m.f.s is equal to the sum of the p.d.s.

Review questions

1. a) State Kirchoff's junction rule.

   b) Explain why it is a consequence of the conservation of charge.

   c) In the following circuit, find the current at points A, B, C and D.

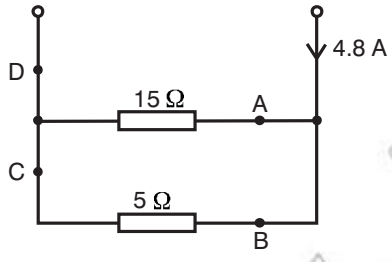


Figure 5.21

2. The voltmeter in this circuit has an infinite resistance.

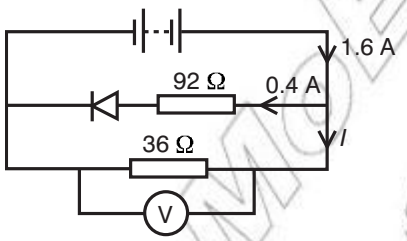


Figure 5.22

   Calculate

   a) the current in the 36 Ω resistor

   b) the reading on the voltmeter.

3. Find the values of the ammeter and voltmeter readings in this circuit. Assume that the ammeter and cell have negligible internal resistance and that the voltmeter has an infinite resistance.

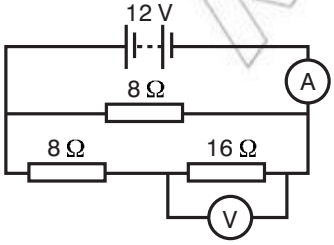


Figure 5.23

© MoE-FDR Ethiopia Not to be republished

5.3 Measuring instruments

By the end of this section you should be able to:

- Describe how a galvanometer can be modified to measure a wide range of currents and potential differences.
- Describe how shunt resistors are used to measure a wide range of currents and p.d.
- Calculate shunt and multiplier value for use with a meter to give different current and voltage ranges.
- Solve problems in which a meter resistance is involved.
- Identify and appropriately use equipment for measuring potential difference, electrical current and resistance (e.g. use multimeters and a galvanometer to make various measurements in an electrical circuit, use an oscilloscope to show the characteristics of the electrical current).

How a galvanometer can be modified using shunt resistors to measure a wide range of currents and p.d.s

In Grade 10, you learnt that the greater the current flowing around the coil of an electric motor, the more strongly it will try to turn. This suggests a way to measure the size of a current: let it flow through a motor, and make the coil try to turn while it is held back by a spring. The bigger the current, the further the coil will manage to stretch the spring.

This is the basis of the moving-coil galvanometer. The coil of the instrument is drawn in Figure 5.24(a). The current can be fed into the coil and out again via the hairsprings at top and bottom; no commutator is needed in this case because the rotation of the coil is restricted to just a fraction of a turn.

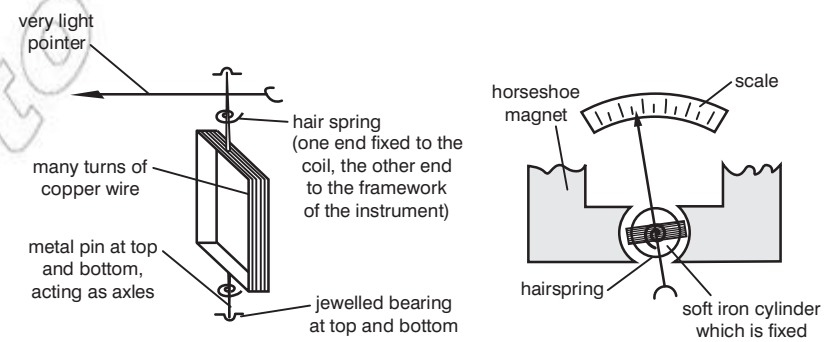


Figure 5.24 The moving-coil galvanometer

Figure 5.24(b) shows a view of the complete arrangement from above. The coil can rotate inside the gap of a steel horse-shoe magnet whose pole pieces are curved. The soft iron cylinder which sits in the middle of the coil (but does not rotate with it) itself gets turned into a magnet because of the presence of the permanent magnet; one of its effects is to increase the strength of the field within the gap.

© MoE Not to be re

Its other effect is to give the instrument a linear scale. In the gap there is a radial field (think of how a small compass would set at that point), so as the coil rotates within the gap it always stays along the field lines. The 'cos θ' term does not appear in the torque, so the torque remains proportional to the current. (You learnt about the torque in a magnetic field in Grade 10.)

A galvanometer thus measures an electric current, 'galvanism' being an old name for current electricity. The greater the current round the coil, the more marked the motor effect is and the further the hairsprings are wound up.

A typical instrument is so sensitive that its pointer will be moved to the end of the scale by a current of perhaps 5×10^{-3} A; we say that it has a full-scale deflection of 5 mA. Even though copper is used for the windings of its coil, it consists of such a long length of so very thin wire that it may have a resistance as high as 50 ohms or more.

An ammeter has a very low resistance and is placed in series in a circuit, as shown in Figure 5.25.

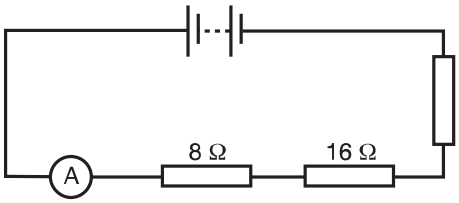


Figure 5.25 *An ammeter in series with resistors*

The basic galvanometer described can be converted into an ammeter by adding a low resistance 'shunt', which is usually fitted inside the casing of the instrument and consists of a short length of quite thick wire (see Figure 5.26). Most of the current takes this low resistance shunt route, and only a tiny proportion trickles through the coil to rotate the pointer.

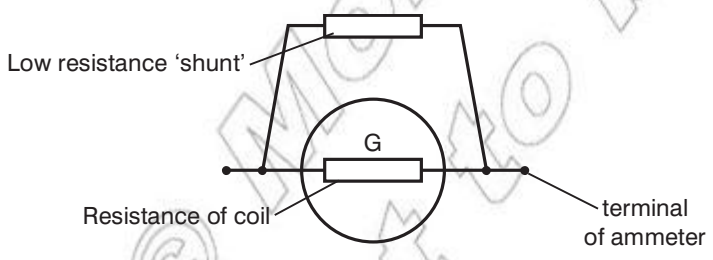


Figure 5.26 *Conversion to an ammeter*

A variety of different range settings can be achieved by varying the shunt resistance so that the amount of current that goes through the shunt varies. The table overleaf shows readings for an instrument that reads to full-scale deflection of 1 mA, 10 mA, 100 mA and 1000 mA. The coil in such a meter will always read to a maximum of 1 mA.

© MoE FDR Ethiopia Not to be republished

Range (mA)	I_{coil} up to	I_{shunt} up to	Fraction in coil
0–1	1 mA	0	1
0–10	1 mA	9 mA	$\frac{1}{9}$
0–100	1 mA	99 mA	$\frac{1}{99}$
0–1000	1 mA	999 mA	$\frac{1}{999}$

A voltmeter has a very high resistance and is placed in parallel with the component. You can convert a galvanometer to be a voltmeter by adding a large resistance in series with the meter, as shown in Figure 5.27.

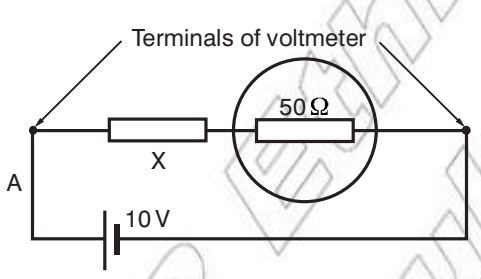


***Figure 5.27** Conversion to a voltmeter*

Calculating shunt and multiplier values for use with a meter to give different current and voltage ranges

You can calculate the value of the shunt resistor that is required to convert a galvanometer to both an ammeter and a voltmeter, as shown in the following worked examples.

Worked example 5.15

A galvanometer of full-scale deflection 5 mA is to be converted into a 0–10 A ammeter. If its coil has a resistance of 50 Ω, what value shunt must be fitted?

Draw the circuit with a current of 10 A flowing (Figure 5.28).

Under these circumstances we want the pointer of the meter just to reach the end of its scale, and this will mean 5 mA (0.005 A) flowing through it.

Use Kirchoff's junction rule to work out the current that must go through the shunt.

10 – 0.005 = 9.995 A

The p.d. between X and Y (V_{xy}) must be sufficient to drive 0.005 A through the 50 Ω of the coil. Thus $V_{xy} = IR = 0.005 \times 50 = 0.25$ V.

Now this p.d. is also across the shunt, so to work out R we must ask what size resistor is needed in order that, with a p.d. of 0.25 V across it, a current of 9.995 A will flow through it.

This gives $R = \frac{V}{I} = \frac{0.25}{9.995} = 0.025\ \Omega$

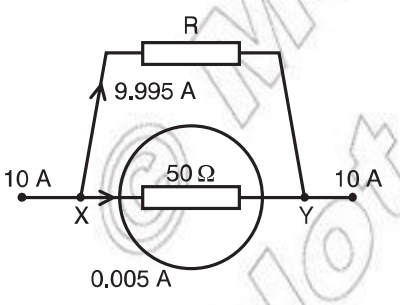


Figure 5.28

Worked example 5.16

A galvanometer of resistance 50 Ω and full-scale deflection 5 mA is to be made into a 0–10 V voltmeter. How can this be done?

Imagine that the voltmeter is to be used to measure a 10 V battery. If the pointer of the galvanometer is just to reach the end of its scale when connected to the battery, a current of 0.005 A must now flow through it (Figure 5.29).

For this current to be drawn from the 10 V battery, the total resistance of the whole circuit must be given by:

$R = \frac{V}{I} = \frac{10}{0.005} = 2000\ \Omega.$

The galvanometer already provides 50 Ω of this, so X = 2000 – 50 = 1950 Ω.

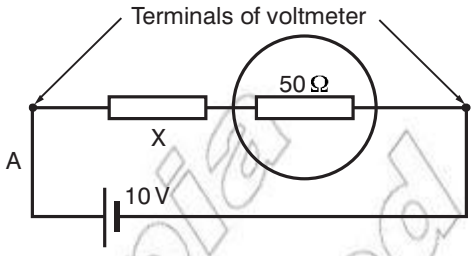


Figure 5.29

Activity 5.8: Converting a galvanometer to an ammeter and voltmeter

Work in a small group. Use the information in this section to build your own ammeter and voltmeter from a basic galvanometer (remember that you built an electric motor in Grade 10).

Solving problems involving a meter resistance

There are circumstances in which the resistance of a measuring meter needs to be taken into account in circuit calculations. The following examples show you how this is done.

Worked example 5.17

An ammeter of resistance R_3 is added to the circuit shown in Figure 5.30.

Before the ammeter was added, the current in the circuit was 0.03 A. When the ammeter is added, it reads 0.02 A. Calculate the resistance of the ammeter.

p.d. (V)	I (A)	R (Ω)
12	0.02	$400 + R_3$

Use Ohm's law $V = IR$ so

$R = \frac{V}{I}$

$400 + R_3 = \frac{12}{0.02}$

$R_3 = 600 - 400$

$= 200\ \Omega$

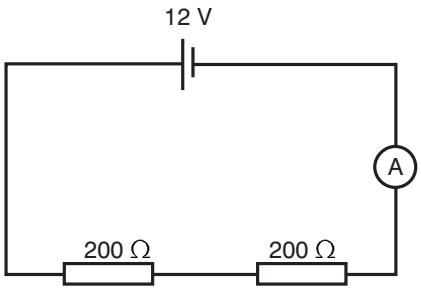


Figure 5.30

Worked example 5.18

When a voltmeter is added to the circuit shown in Figure 5.31, it acts as a parallel resistance of resistance R_3. The supply has negligible internal resistance.

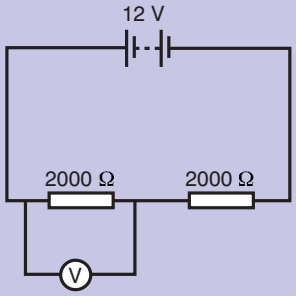


Figure 5.31

a) Show that the p.d. across one 2000 Ω resistor before the voltmeter is added is 6 V.

b) The voltmeter reads 4 V when it is in the circuit. Calculate its resistance R_3.

a) Without the voltmeter in the circuit the total resistance is 4000 Ω and the total p.d. is 12 V.

p.d. (V)	I (A)	R (Ω)
12	?	4000

Using Ohm's law

$$I = \frac{V}{R}$$
$$= \frac{12}{4000}$$
$$= 3 \times 10^{-3} \text{ A}$$

This is the current that flows through a 2000 Ω resistor.

The p.d. across this resistor is therefore = $3 \times 10^{-3} \times 2000 = 6$ V

b) When the resistance of the voltmeter is taken into account, the circuit can be drawn as shown in Figure 5.32.

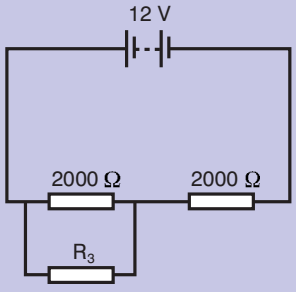


Figure 5.32

© MoE Not for

The resistance of the parallel combination is therefore

$$\frac{2000R_3}{2000 + R_3}\ \Omega$$

Using Kirchoff's loop rule, if the p.d. across the parallel combination is 4 V, then the p.d. across the single resistor must be 8 V.

This means that the current flowing through the circuit is now

$$= \frac{8}{2000}$$

$$= 4 \times 10^{-3}\ \text{A}$$

p.d. (V)	I (A)	R (Ω)
4	4×10^{-3}	$\frac{2000R_3}{2000 + R_3}$

Use Ohm's law $R = \frac{I}{R}$

$$\frac{2000R_3}{2000 + R_3} = \frac{4}{4 \times 10^{-3}}$$

$$= 1000$$

$$2000R_3 = 1000(2000 + R_3)$$

$$2000R_3 = 2\ 000\ 000 + 1000R_3$$

$$R_3 = 2000\ \Omega$$

Equipment for measuring potential difference, electrical current and resistance

Different instruments are used for different kinds of measurements in electrical circuits. So far in this unit you have used ammeters to measure current, and voltmeters to measure p.d. You can also use these instruments to find out the resistance of a component by plotting its V–I characteristics, as you did in Activity 5.5.

Activity 5.9: Using electrical measuring instruments

Work in a small group to produce a poster that summarises all you have learnt about using measuring instruments in electrical circuits.

Think about this...

In Unit 7 you will study alternating current. You can use an instrument called an oscilloscope to show the characteristics of this type of current.

© Not to be republished

Summary

In this section you have learnt that:

- A galvanometer can be modified to measure a wide range of currents and potential differences using shunt resistors.
- A range of equipment can be used for measuring potential difference, electrical current and resistance (e.g. multimeters and a galvanometer can be used to make various measurements in an electrical circuit).

Review questions

1. A galvanometer has a resistance of 40 Ω and is of 3 mA full-scale deflection. How would you modify it to act as a 0–5 A ammeter?
2. The galvanometer described in question 1 is to be converted into a 0–5 V voltmeter.
   a) When the voltmeter is connected to a 5 V supply, how great a current will need to flow through it?
   b) What must the resistance be between the terminals of the voltmeter for that to happen?
3. An ammeter of resistance R is added to the circuit shown in Figure 5.33.

   Before the ammeter was added, the current in the circuit was 0.02 A. When the ammeter is added, it reads 0.015 A. Calculate the resistance of the ammeter.
4. How should a) an ammeter b) a voltmeter be connected in a circuit to function correctly?

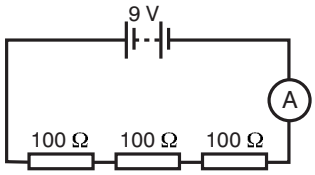


Figure 5.33

5.4 The Wheatstone bridge and the potentiometer

By the end of this section you should be able to:

- Explain the principle of the Wheatstone bridge and solve problems involving it.
- Explain the principle of the potentiometer and how it can be used for measurement of e.m.f., p.d., resistance and current.
- Solve problems involving potentiometer circuits.

© MoE, FDRE Ethiopia
Not to be republished

The Wheatstone bridge

The basic bridge circuit is shown in Figure 5.34. The fundamental concept of the Wheatstone bridge is that two voltage, or potential, dividers in the same circuit are both supplied by the same input, as shown in Figure 5.34. The circuit output is taken from both voltage divider outputs, as shown.

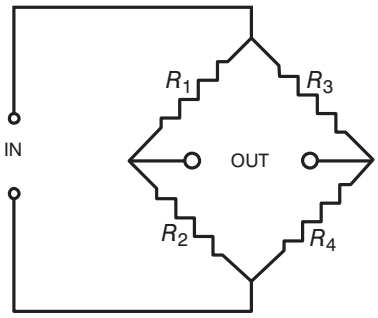


Figure 5.34 *A Wheatstone bridge circuit*

In its classic form, a galvanometer is connected between the output terminals, and is used to monitor the current flowing from one voltage divider to the other. If the two voltage dividers have exactly the same ratio ($\frac{R_1}{R_2} = \frac{R_3}{R_4}$), then the bridge is said to be *balanced* and no current flows in either direction through the galvanometer.

If one of the resistors changes even a little bit in value, the bridge will become unbalanced and current will flow through the galvanometer. Thus, the galvanometer becomes a very sensitive indicator of the balance condition.

In its basic application, a d.c. voltage (E) is applied to the Wheatstone bridge, and a galvanometer (G) is used to monitor the balance condition. The values of R_1 and R_3 are precisely known, but do not have to be identical. R_2 is a calibrated variable resistance, the current value of which may be read from a dial or scale.

An unknown resistor, R_X, is connected as the fourth side of the circuit, as shown in Figure 5.35, and power is applied. R_2 is adjusted until the galvanometer, G, reads zero current. At this point, $R_X = R_2 \times \frac{R_3}{R_1}$.

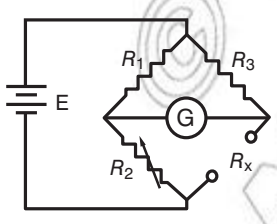


Figure 5.35

DID YOU KNOW?

The circuit we now know as the Wheatstone bridge was actually first described by Samuel Hunter Christie (1784–1865) in 1833.

However, Sir Charles Wheatstone invented many uses for this circuit once he found the description in 1843. As a result, this circuit is known generally as the Wheatstone bridge. It is not possible to cover all of the practical variations and applications of the Wheatstone bridge. Sir Charles Wheatstone invented many uses himself, and others have been developed since that time. One very common application in industry today is to monitor sensor devices such as strain gauges. Such devices change their internal resistance according to the specific level of strain (or pressure, temperature, etc.), and serve as the unknown resistor R_X. However, instead of trying to constantly adjust R_2 to balance the circuit, the galvanometer is replaced by a circuit that can be calibrated to record the degree of imbalance in the bridge as the value of strain or other condition being applied to the sensor.

To this day, the Wheatstone bridge remains the most sensitive and accurate method for precisely measuring resistance values.

© Not to be republished

Activity 5.10: Using a Wheatstone bridge (1)

Set up the circuit shown in worked example 5.19 and check that the galvanometer reading is zero when R_x is 200 Ω.

Activity 5.11: Using a Wheatstone bridge (2)

Use a Wheatstone bridge circuit to find an unknown resistance and determine the specific resistance (resistance per unit length) of a wire.

Activity 5.12: Applications of a Wheatstone bridge

In a small group, research some applications of a Wheatstone bridge. Present your findings to the rest of the class in a form of your choice.

Worked example 5.19

A Wheatstone bridge circuit is set up as shown in Figure 5.36. It is balanced. What is the resistance, R_x?

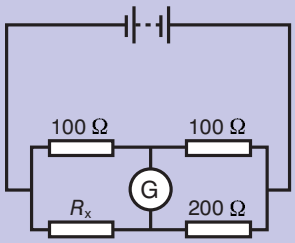


Figure 5.36

In this circuit, the values of the resistances are as follows

R_1	R_2 (Ω)	R_3 (Ω)	R_x (Ω)
100	100	200	?

Use $R_x = \frac{R_2 \times R_3}{R_1} = \frac{100 \times 200}{100}$

$= 200\ \Omega$

The potentiometer

A potentiometer has a sliding contact and acts as an adjustable potential divider. The basic circuit is shown in Figure 5.37.

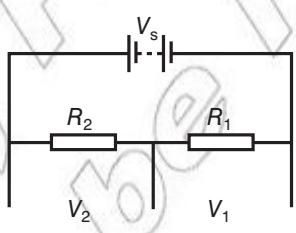


Figure 5.37

You know that the current through the two series resistors R_1 and R_2 will be the same. By adjusting the position of the sliding contact, you adjust the values of R_1 and R_2, and hence the value of the potential difference V_1 and V_2.

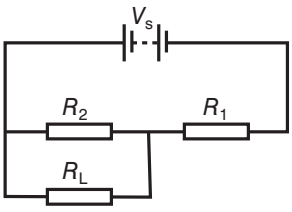


Figure 5.38

If we now put a load resistance R_L in parallel with R_2 as shown in Figure 5.38, then the potential difference V_L across R_L can be calculated using the equation

© MoE Not to be reproduced

$$V_L = \frac{R_2 R_L}{R_1 R_L + R_2 R_L + R_1 R_2} V_S$$

However, if R_L is large in comparison with R_1 and R_2 (as it would be in a practical application such as the input to an operational amplifier) then this equation can be simplified to

$$V_L = \frac{R_2}{R_1 + R_2} V_S$$

Similarly, if R_L is in parallel with R_1, the potential difference V_L is given by

$$V_L = \frac{R_1}{R_1 + R_2} V_S$$

You can see that the supply voltage (e.m.f.) is divided by this circuit in proportion to the values of R_1 and R_2.

Activity 5.13: How could a potential divider circuit be used to measure e.m.f.?

In a small group, use the above information to work out how a potential divider circuit could be used to measure e.m.f. What measurements would need to be taken? What calculations would need to be done?

Activity 5.14: Using a potential divider circuit to compare the e.m.f. of two cells

Devise and carry out a way of comparing the e.m.f.s of two cells using a potentiometer.

Activity 5.15: Using a potential divider circuit to find the internal resistance of a cell

Devise and carry out a way of finding the internal resistance of a cell using a potentiometer.

Activity 5.16: How could a potential divider by used to measure current?

In a small group, use the above information to work out how a potential divider circuit could be used to measure current. What measurements would need to be taken? What calculations would need to be done?

Activity 5.17: Applications of a potentiometer

In a small group, research practical applications of potentiometers. For example, how are they used in audio control? Present your findings to the rest of your class in a form of your choice.

© Ethiopia published

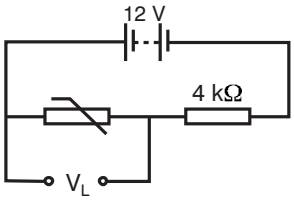


Figure 5.39

Worked example 5.20

A thermistor is connected in circuit shown in Figure 5.39.

The resistance of the thermistor varies with temperature. It has a resistance of 12 kΩ at a temperature of 0 °C and a resistance of 0.25 kΩ at a temperature of 25 °C.

Find the output p.d. (V_L) at a) 0 °C and b) 25 °C.

Use $V_L = \frac{R_2}{R_1 + R_2} V_S$

R_2 = resistance of thermistor

$R_1 = 4\ \text{k}\Omega$

$V_S = 12\ \text{V}$

a) $V_L = \frac{12}{12 + 4} \times 12$

$= 9\ \text{V}$

b) $V_L = \frac{0.25}{0.25 + 4} \times 12$

$= 0.706\ \text{V}$

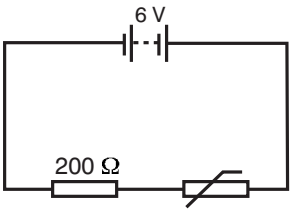


Figure 5.40

Worked example 5.21

A thermistor is used in a circuit shown in Figure 5.40.

At a temperature of 60 °C the resistance of the thermistor is 100 Ω. At this temperature:

a) what is the p.d. across the thermistor?

b) what is the current flowing in the circuit?

a) Use

$V_T \frac{R_2}{R_1 + R_2} V_S$

R_2 = resistance of thermistor

$R_1 = 200\ \Omega$

$V_S = 6\ \text{V}$

a) $V_T = \frac{100}{200 + 100} \times 6$

$= 2\ \text{V}$

b) To find current in circuit, use Ohm's law on thermistor:

$I = \frac{V}{R}$

$= \frac{2}{100}$

$= 0.02\ \text{A}$

© MoE Not to

Summary

In this section you have learnt that:

- The principle of the Wheatstone bridge circuit is that an unknown resistance can be determined by placing it in the circuit along with three known resistances and then finding the point at which there is no current flowing through a galvanometer because the two potential dividers have exactly the same ratio.
- The principle of the potentiometer is that the supply voltage is divided between the resistors in the circuit in the ratio of the values of the resistances. The formula is as follows:

$$V_L = \frac{R_2}{R_1 + R_2} V_S$$

  where V_L is the potential difference across the load, V_S is the supply voltage and R_1 and R_2 are the resistances used in the circuit.
- A potentiometer can be used for measurement of e.m.f., since once the p.d. across each resistor has been found by measurement, Kirchoff's loop rule can be applied to sum these and thus find the supply voltage.
- A potentiometer can be used to determine p.d. because the p.d. across one of the resistors can be calculated if the values of the two resistances and the supply voltage are known.
- If the p.d. across an unknown resistance in a potentiometer circuit is measured, then provided that the supply voltage and the value of the other resistance are known, the unknown resistance can be calculated.
- A current in a potentiometer circuit can be determined by measuring the p.d. across a known resistance and then using Ohm's law to calculate the current.

Review questions

1. Explain the principle of the Wheatstone Bridge circuit.
2. In the balanced Wheatstone Bridge circuit shown in Figure 5.41, the known resistances are as indicated. Find the value of R.

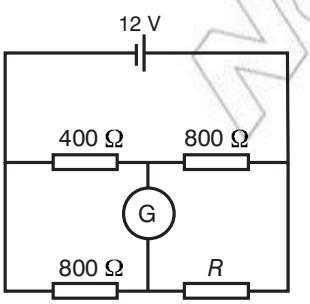


Figure 5.41

Ethiopia Not to be published

3. Explain how a potentiometer circuit can be used to determine p.d.
4. The circuit shown in Figure 5.42 is a touch sensor.

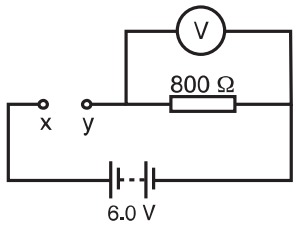


Figure 5.42

When a finger is placed over the contacts *X* and *Y*, the voltmeter reads 3.6 V because of the electrical resistance of the skin. What is the electrical resistance in kΩ between the two contacts?

End of unit questions

1. A small radio receiver uses a battery that delivers a constant current of 50 mA for 6 hours. Calculate the total charge delivered by the battery.
2. A student connects the ends of a pencil 'lead' to a 6.0 V supply and measures the current to be 8.6 A. The 'lead' is a rod of graphite of length 7.5 cm and diameter 1.4 mm.
   a) Calculate the resistance of the pencil.
   b) Use this data to calculate the resistivity of graphite.
3. Find the approximate current density when an electric field of 6 V/m is applied to a tungsten filament. The resistivity of tungsten is 5.6×10^{-8} Ω m.
4. Copper contains 8.0×10^{28} free electrons per m^3. A copper wire of cross-sectional area 1.5×10^{-6} m^2 carries a current of 0.5 A. Calculate the drift velocity, in m/s, of the free electrons in the wire. (The elementary charge is 1.6×10^{-19} C.)
5. a) State Kirchoff's loop rule.
   b) Explain why it is a consequence of the conservation of energy.
   c) In the circuit in Figure 5.43, find the current in each of the identical resistors points A, B, C, D and E.

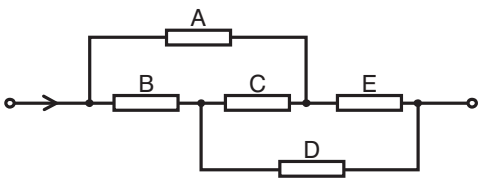


Figure 5.43

6. When a circuit is connected as shown in Figure 5.44, the current is 0.25 A.

© MoE, FDRE Ethiopia
Not to be reproduced

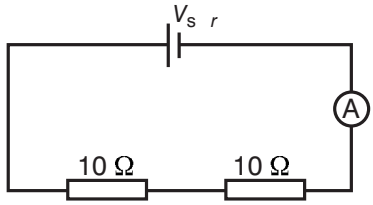


Figure 5.44

When the same components are connected in the circuit as shown in Figure 5.45, the current is 0.67 A.

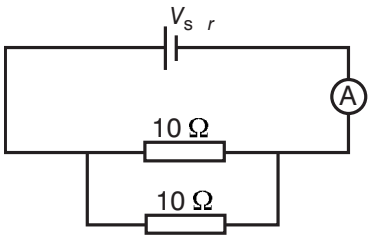


Figure 5.45

Calculate

a) the internal resistance of the cell
b) the e.m.f. of the cell.

7. Explain the basis of a moving coil galvanometer.
8. How can different range settings on a galvanometer be achieved?
9. A galvanometer has a resistance of 60 Ω and is of 5 mA full scale deflection. How would you modify it so as to get a 0–10 A ammeter?
10. The galvanometer described in question 7 is to be converted into a 0–10 V voltmeter.
    a) When the voltmeter is connected to a 10 V supply, how great a current will need to flow through it?
    b) What must the resistance be between the terminals of the voltmeter for that to happen?
11. Explain how the circuit shown below works.

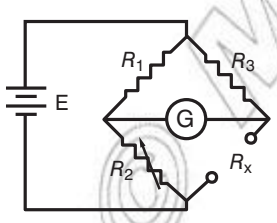


Figure 5.46

12. Draw a diagram of a basic potentiometer circuit.
13. Explain the principle of the potentiometer circuit.
14. Explain how a potentiometer circuit can be used to compare two e.m.f.s.
15. Draw a circuit to show how a thermistor, a voltmeter, a resistor and a cell may be used to control the output from a heater. Explain the operation of your circuit.

© MoE-FDR Ethiopia Not to be republished

Magnetism

Unit 6

Contents

Section	Learning competencies
6.1 Concepts of a magnetic field (page 235)	• Define magnetic field. • State the properties of magnetic field lines. • Describe the properties, including the three-dimensional nature, of magnetic fields.
6.2 The Earth and magnetic fields (page 238)	• Describe magnetic properties of matter. • Distinguish between the terms diamagnetic, paramagnetic and ferromagnetic materials. • Describe the causes of the Earth's magnetism.
6.3 Motion of charged particles in a magnetic field (page 240)	• Describe the motion of a charged particle in a magnetic field. • Identify a moving charge sets up a magnetic field. • Use the equation $F = qv \times B$ to determine the magnitude and direction of the force. • Use the expression for the force on a charged particle in a magnetic field. • Solve problems on the motion of charged particles in electric and magnetic fields. • Describe the path if $\theta \neq 90°$. • Describe J.J. Thompson's experiment of charge to mass ratio. • Determine the value of charge mass ratio for this specific experiment.
6.4 Magnetic force on current-carrying conductors (long, straight, circular loop) (page 247)	• Derive the expression $F = I(l \times B)$. • Use the expression for the force on a current-carrying conductor in a magnetic field. • Determine the magnitude and direction of torque acting on a current loop. • Define magnetic dipole moment. • Describe the working mechanism of a direct motor. • Describe and illustrate the magnetic field produced by an electric current in along straight conductor. • Calculate the magnetic field strength of a straight current carrying wire. • Analyse and predict using the right hand rule the direction of the magnetic field produced when electric current flows through a long straight conductor. • State the Biot–Savart law. • Apply and use the Biot–Savart law to determine the expression for magnetic field strength of a current element.

Contents

Section	Learning competencies
6.5 Ampere's law and its application (page 256)	• State Ampere's law and use it in solving problems. • Describe and illustrate the magnetic field produced in a solenoid and predict its direction using the right hand rule.
6.6 Earth's magnetism (page 260)	• Determine the horizontal component of the Earth's magnetic field at a location. • Resolve the horizontal and vertical components of the Earth's magnetic field. • Describe how a tangent galvanometer works.

6.1 Concepts of a magnetic field

By the end of this section you should be able to:

- Define magnetic field.
- State the properties of magnetic field lines.
- Describe the properties, including the three-dimensional nature, of magnetic fields.

Magnetic fields

You learnt about magnets and **magnetic fields** in Grade 10. You know that a magnet has two poles, which we label 'north' and 'south' (see Figure 6.1).



Figure 6.1

A magnetic field is a region where a magnet exerts a force. **Magnetic field lines** (also called magnetic flux lines), which show where the magnetic field has the same strength, point from the north pole of a magnet to the south pole of the magnet, as shown in Figure 6.2.

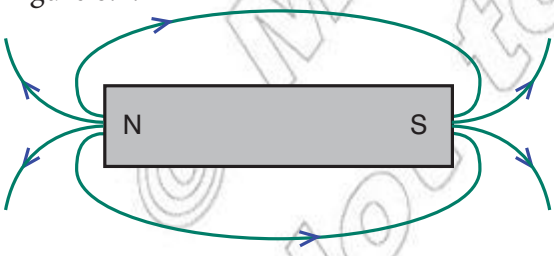


Figure 6.2

A magnetic field is a vector quantity as it has both magnitude (which depends on the strength of the magnet which produced it) and direction. The stronger the magnetic field, the closer the flux lines. Using this idea, we sometimes refer to the magnetic field strength as the magnetic flux density. The symbol for **magnetic flux density** is B and the unit of measurement is the tesla (T).

KEY WORDS

magnetic field *a region where a magnet exerts a force*

magnetic field lines *show where the magnetic field has the same strength*

magnetic flux density *a measure of the strength of the magnetic field – shown visually by how close the lines of flux are to each other*

© Not to be republished

Activity 6.1: Demonstrating the magnetic flux lines around a magnet

In a small group, devise a method for showing the magnetic flux lines around a bar magnet
a) in two dimensions
b) in three dimensions.
(Hint: Think about how you could use a clear plastic bottle, liquid glycerine, iron filings and a bar magnet to show the field in three dimensions.)

Repeat this activity for a different magnet (e.g. a horseshoe magnet) of your choice.

The strength of a magnetic field is also indicated by the quantity of flux (Φ) through any given area. Flux is measured in Webers (Wb). To find the flux for a particular region you multiply the area of the region by the component of flux density perpendicular to the area:

$\Phi = B\sin\theta \times A$

Worked example 6.1

The bar magnet in Figure 6.3 causes a magnetic field with a strength of 30 mT at an angle of 75° to the region of area A, how much flux will be contained by this region if the area is 5 cm^2?

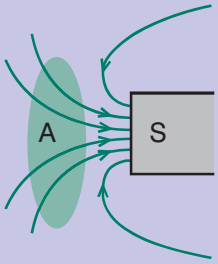


Figure 6.3

Φ (Wb)	B (T)	sin θ	A (m^2)
?	30×10^{-3}	0.966	5×10^{-4}

Use $\Phi = B\sin\theta \times A$

$= 30 \times 10^{-3} \times 0.966 \times 5 \times 10^{-4}$

$= 1.45 \times 10^{-5}$ Wb

© MoE Not to be republished

Summary

In this section you have learnt that:

- A magnetic field is a region where a magnet exerts a force. Magnetic field lines, (also called magnetic flux lines) which show where the magnetic field has the same strength, point from the north pole of a magnet to the south pole of the magnet.
- A magnetic field is a vector quantity as it has both magnitude (which depends on the strength of the magnet which produced it) and direction. The stronger the magnetic field, the closer the flux lines. Using this idea, we sometimes refer to the magnetic field strength as the magnetic flux density. The symbol for magnetic flux density is B and the unit of measurement is the tesla (T).
- The strength of a magnetic field is also indicated by the quantity of flux (Φ) through any given area. Flux is measured in Webers (Wb). To find the flux for a particular region you multiply the area of the region by the component of flux density perpendicular to the area:
  $\Phi = B\sin\theta \times A$

Review questions

1. What is a magnetic field?
2. Copy the diagrams and draw in the magnetic flux lines for each magnet.

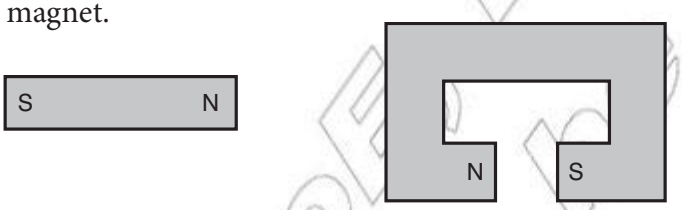


Figure 6.4

3. How could you demonstrate the three-dimensional nature of the magnetic field around a bar magnet?
4. The bar magnet in Figure 6.5 causes a magnetic field with a strength of 20 mT at an angle of 60° to the region of area A, how much flux will be contained by this region if the area is 10 cm²?

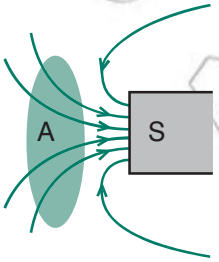


Figure 6.5

© MoE FDRE Not to be republished

KEY WORDS

magnetism *describes how the atoms of materials respond to a magnetic field*

diamagnetism *the tendency of a material to oppose an applied magnetic field*

paramagnetic *materials have unpaired electrons, which will tend to align themselves in the same direction as the applied magnetic field, thus reinforcing it*

ferromagnetic *materials have unpaired electrons, which will align with the applied magnetic field and parallel to each other. They keep this alignment even when the applied field is removed.*

lodestone *a naturally magnetised piece of the mineral magnetite*

DID YOU KNOW?

A **lodestone** or **loadstone** is a naturally magnetised piece of the mineral magnetite. They are naturally occurring magnets that attract pieces of iron. Ancient people first discovered the property of magnetism in lodestone. The earliest written reference to magnetism occurs in a book from the 4th Century BCE in ancient China. Pieces of lodestone, suspended so they could turn, were the first magnetic compasses, and their importance to early navigation is indicated by the name lodestone, which in Middle English means 'course stone' or 'leading stone'.

6.2 The Earth and magnetic fields

By the end of this section you should be able to:

- Describe magnetic properties of matter.
- Distinguish between the terms diamagnetic, paramagnetic and ferromagnetic materials.
- Describe the causes of the Earth's magnetism.

Magnetic properties of matter

We use the term **magnetism** to describe how the atoms of materials respond to a magnetic field. **Diamagnetism** is a property of all materials. It is the tendency of a material to oppose an applied magnetic field. Some materials will, however, reinforce a magnetic field because they have unpaired electrons, which will tend to align themselves in the same direction as the applied magnetic field, as shown in Figure 6.6. These are known as **paramagnetic** materials.

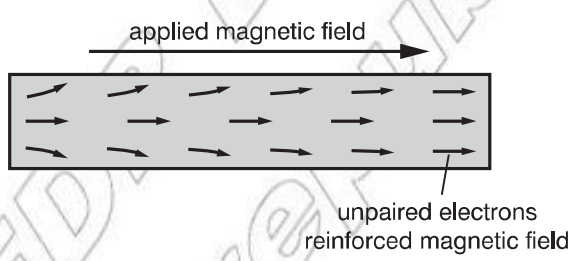


***Figure 6.6** The alignment of unpaired electrons*

Ferromagnetic materials have unpaired electrons. In addition to the tendency of these electrons to align themselves in the same direction as an applied magnetic field, they will also align themselves so that they are parallel to each other. This means that, even when the applied field is removed, the electrons in the material maintain a parallel orientation. Examples of ferromagnetic materials are nickel, iron, cobalt and their alloys.

The causes of the Earth's magnetism

The Earth can be thought of as a huge magnet. The geographic north pole is the south pole of the magnet, and the geographic south pole is the north pole of the magnet. Hence, the needle on a compass will be attracted to the geographic north pole (see diagram).

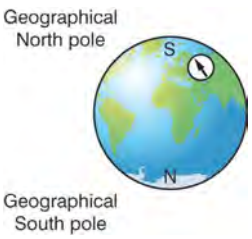


***Figure 6.7** The poles of the Earth's magnetic field*

There is no simple answer to the question 'where does the Earth's magnetism come from?'. In Grade 10, you learnt that magnetic fields surround electric currents. We can form a theory that circulating electric currents in the molten iron core of the Earth produce the magnetic field. We do not know how this 'dynamo effect' works in detail, but the rotation of the Earth plays a part in generating the currents that are presumed to be the source of the magnetic field. You will discuss the horizontal and vertical components of the Earth's magnetic field in more detail in section 6.6.

DID YOU KNOW?

The properties of magnets and the dry compass were discovered in 1282 by a Yemeni physicist, astronomer and geographer, Al-Ashraf.

Activity 6.2: Plotting the combined magnetic field of the Earth and a bar magnet

Place a bar magnet with its north pole facing geographic south on large sheet of paper. Use a compass to plot the combined magnetic field of the Earth and this magnet. Find any neutral points (that is, points where there is no field).

Summary

In this section you have learnt that:

- The term **magnetism** describes how the atoms of materials respond to a magnetic field.
- **Diamagnetism** is a property of all materials. It is the tendency of a material to oppose an applied magnetic field.
- **Paramagnetic** materials reinforce a magnetic field because they have unpaired electrons which will tend to align themselves in the same direction as the applied magnetic field.
- **Ferromagnetic** materials have unpaired electrons. In addition to the tendency of these electrons to align themselves in the same direction as an applied magnetic field, they will also align themselves so that they are parallel to each other. This means that, even when the applied field is removed, the electrons in the material maintain a parallel orientation.
- We can form a theory that circulating electric currents in the molten iron core of the Earth produce the magnetic field. We do not know how this 'dynamo effect' works in detail, but the rotation of the Earth plays a part in generating the currents which are presumed to be the source of the magnetic field.

Review questions

1. Explain the difference between diamagnetic, paramagnetic and ferromagnetic materials.
2. Outline a theory that can explain the Earth's magnetic field.

6.3 Motion of charged particles in a magnetic field

By the end of this section you should be able to:

- Describe the motion of a charged particle in a magnetic field.
- Identify a moving charge sets up a magnetic field.
- Use the equation $F = qv \times B$ to determine the magnitude and direction of the force.
- Use the expression for the force on a charged particle in a magnetic field.
- Solve problems on the motion of charged particles in electric and magnetic fields.
- Describe the path if $\theta \neq 90°$.
- Describe J.J. Thompson's experiment of charge to mass ratio.
- Determine the value of charge mass ratio for this specific experiment.

The motion of a charged particle in a magnetic field

A moving charged particle creates a magnetic field. A charged particle moving in a magnetic field will create a force, $\vec{F}$. The magnitude of this force depends on:

- the speed of the particle, $\vec{v}$
- the strength of the magnetic field, $\vec{B}$.

The force can be calculated using the vector cross product

$$\vec{F} = q(\vec{v} \times \vec{B})$$

From the definition of the vector cross product, we can say that the magnitude of the force is

$$F = qvB\sin\theta$$

where θ is the angle between $\vec{v}$ and $\vec{B}$. We can find the direction of the force by using the right hand rule, as shown in Figure 6.8.

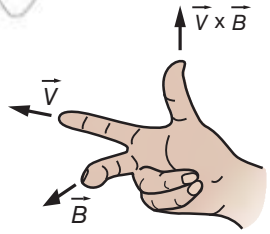


***Figure 6.8** The right hand rule for magnetic force*

© MoE-TDC Not to be reproduced

Worked example 6.2

a) Find the size of the force felt by an electron travelling perpendicular to the Earth's magnetic field at 500 m/s. (The charge on an electron is 1.6×10^{-19} C and the magnitude of the Earth's magnetic field is 5×10^{-5} T.)

b) In what direction will the force act?

a)

F (N)	q (C)	v (m/s)	B (T)	sin θ
?	1.6×10^{-19}	500	5×10^{-5}	1

Use

$\vec{F} = q(\vec{v} \times \vec{B})$

$F = qvB\sin\theta$

$= 1.6 \times 10^{-19} \times 500 \times 5 \times 10^{-5} \times 1$

$= 4 \times 10^{-21}$ N

b) The force will act in a direction that is perpendicular to both the Earth's magnetic field and the direction in which the electron is travelling.

Worked example 6.3

Find the size of the force felt by an electron travelling at an angle of 30° to the Earth's magnetic field at 500 m/s. (The charge on an electron is 1.6×10^{-19} C and the magnitude of the Earth's magnetic field is 5×10^{-5} T.)

F (N)	q (C)	v (m/s)	B (T)	sin θ
?	1.6×10^{-19}	500	5×10^{-5}	0.5

Use

$\vec{F} = q(\vec{v} \times \vec{B})$

$F = qvB\sin\theta$

$= 1.6 \times 10^{-19} \times 500 \times 5 \times 10^{-5} \times 0.5$

$= 2 \times 10^{-21}$ N

Activity 6.3: Determining the strength of a magnetic field

In a small group, discuss how you could determine the strength of a magnetic field if you had a small test charge whose acceleration you were able to measure.
(Hint: start with Newton's second law.)

Ethiopia Not to be republished

The motion of charged particles in electric and magnetic fields

Particles can move in both magnetic and electric fields. If a particle is simply moving in an electric field, then you know that

$F = Eq$

where F is the force experienced by the particle, E is the strength of the magnetic field and q is the charge on the particle.

There are useful devices that use a combination of electric and magnetic fields. An example is a velocity selector. This device uses a combination of electric and magnetic fields to trap particles moving at different speeds. When the force on a particle as a result of the electric field is the same as the force on the particle as a result of the magnetic field

$F = Eq = qvB\sin\theta$

Hence $v = \frac{E}{B}\sin\theta$

Worked example 6.4

Find the speed of an electron travelling at 90° to the electric and magnetic fields in a velocity selector operating with an electric field of 3.0 kV and a magnetic field of 3.0 T.

v (m/s)	E (V)	B (T)	$\sin\theta$
?	3.0×10^3	3.0	1

Use $v = \frac{E}{B}\sin\theta$

$= \frac{3 \times 10^3}{3}$

$= 1 \times 10^3$ m/s

Activity 6.4: Researching how electric and magnetic fields are used in traditional television and computer screens

In a small group, carry out some research to find out how electric and magnetic fields are used in traditional television and computer screens.

J.J. Thompson's experiment of charge to mass ratio

J.J. Thompson used balanced electric and magnetic fields to measure the charge to mass ratio for an electron. His apparatus is shown in Figure 6.9.

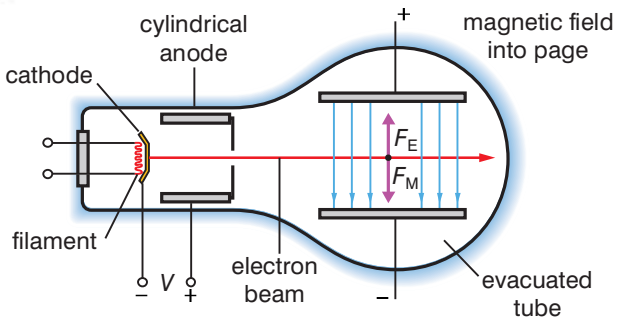


Figure 6.9 *Measuring the change to mass ratio of an election*

In this case θ = 90°. We know that in such circumstances

$$v = \frac{E}{B}$$

We can find another expression for v by using the fact that the electron beam is accelerated by the potential difference between the cathode and the anode. We know that the kinetic energy of the electrons is given by

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

where v is the velocity of an electrons, m is the mass of an electron, q is the charge on an electron and V is the accelerating potential difference.

We can rearrange this equation to

$$v = \sqrt{\frac{2qV}{m}}$$

If we equate this to the expression for v involving E and B, and square both sides, we get

$$\frac{E^2}{B^2} = \frac{2qV}{m}$$

We can rearrange this to get

$$\frac{q}{m} = \frac{E^2}{2VB^2}$$

DID YOU KNOW?

John Joseph Thomson won the Nobel Prize for Physics in 1906. In 1937, his son, George Paget Thomson, won the same prize.

Worked example 6.5

a) Find the charge mass ratio for an electron accelerated through 600 V in a magnetic field of strength 45 mT where the speed of the electron is 1.4×10^7 m/s.

b) What is the percentage difference between your result and the accepted ratio with the values $q = 1.6 \times 10^{-19}$ C and $m = 9.11 \times 10^{-31}$ kg?

a)

v (m/s)	V (V)	q/m (C/kg)
1.4×10^7	600	?

Use

$$v = \sqrt{\frac{2qV}{m}}$$

Square both sides

$$v^2 = \frac{2qV}{m}$$

Rearrange

$$\frac{q}{m} = \frac{v^2}{2V}$$

FDR Ethiopia Unpublished

$$= \frac{(1.4 \times 10^7)^2}{2 \times 600}$$
$$= 1.63 \times 10^{11}\ \text{C/kg}$$

b) Ratio with accepted values $= \frac{1.6 \times 10^{-19}}{9.11 \times 10^{11}}$
$$= 1.76 \times 10^{11}$$

$$\text{Percentage difference} = \frac{1.76 \times 10^{11} - 1.63 \times 10^{11}}{1.76 \times 10^{11}} \times 100$$
$$= 7.4\%$$

Circular motion of particles in magnetic fields

From the right hand rule, you know that the force on a charged particle is always at right angles to the direction of its velocity. The force therefore acts as a centripetal force and so the particle follows a circular path.

You know that $F = qvB$ and $F = \frac{mv^2}{r}$

where q is the charge on the particle, v is the velocity of the particle, B is the strength of the magnetic field and r is the radius of the circular path.

From this we can see that

$$r = \frac{mv}{qB}$$

We can also find the period, T, since

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

The frequency, f, is the inverse of the period so

$$f = \frac{qB}{2\pi m}$$

The angular velocity, ω, is $2\pi f$, so

$$\omega = \frac{qB}{m}$$

The mass spectrometer

A mass spectrometer is a machine that allows chemicals to be separated according to their mass. A simplified diagram of a mass spectrometer is shown in Figure 6.10.

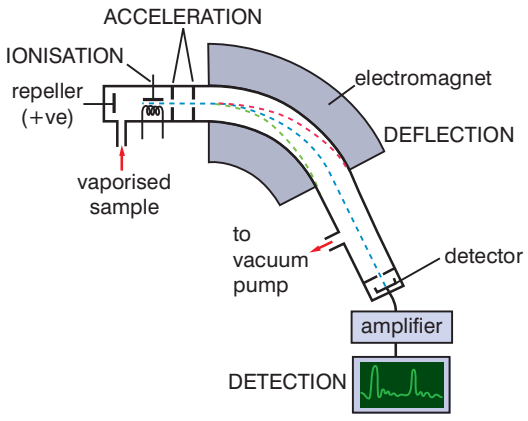


Figure 6.10 *A mass spectrometer*

© MoE Not to be republished

The chemical enters the machine and is ionised (charged). It is then accelerated by an electric field and then its direction is changed when it enters a magnetic field.

In the last section, we learnt that in a magnetic field, a charged particle experiences a force as a result of the magnetic field,

$F = Bqv$

and a centripetal force

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

Particles that follow the central dotted path in Figure 6.9 will reach the detector. For these particles

$F = Bqv = \frac{mv^2}{r}$ where r is the radius of the circular path.

We can rearrange this to $\frac{q}{m} = \frac{v}{Br}$

We can identify the particles by the value of the charge : mass ratio, and the values of B and r are known from the calibration of the machine. So we just need to know the speed of the particles when they entered the electromagnet. From the section on J. J. Thomson's experiment above, we know that

$v = \sqrt{\frac{2qV}{m}}$

If we substitute this into the equation for the charge to mass ratio

$\frac{q}{m} = \frac{\sqrt{2qV}}{Br\sqrt{m}}$

Square both sides

$\frac{q^2}{m^2} = \frac{2qV}{B^2r^2m}$

Thus

$\frac{q}{m} = \frac{2V}{B^2r^2}$

So we can find the mass of a particle using

$m = \frac{B^2r^2q}{2V}$

So by adjusting the accelerating voltage and the strength of the electromagnet (by changing the current through it) we can identify different chemicals in a sample.

© MoE, FDR Ethiopia Not to be republished

Summary

In this section you have learnt that:

- A moving charge sets up a magnetic field.
- A charge moving in a magnetic field will travel in a circular path whose radius, r, is given by $r = \frac{mv}{qB}$ where m is the mass of the particle, v is the speed of the particle, q is the charge on the particle and B is the strength of the magnetic field.
- The equation $F = qv \times B$ is used to determine the magnitude and direction of the force.
- If $\theta \neq 90°$ then the force on the particle will be reduced by a factor $\sin \theta$ from its maximum value, which occurs when the particle is travelling perpendicular to the magnetic field.
- J.J. Thompson's experiment to find charge to mass ratio is based on applying equal forces from an electric field and from a magnetic field to a charged particle. Two expressions for the velocity of the particle are found, one from the balanced forces and the other from the kinetic energy of the particle. These two expressions are equated and a value for the charge to mass ratio is found to be $\frac{q}{m} = \frac{E^2}{2VB^2}$

Review questions

1. Derive an expression for the radius, r, of the circular path of a particle of charge q and of mass m moving at speed v in a magnetic field of strength B.
2. Find the size of the force felt by an electron travelling at an angle of 50° to the Earth's magnetic field at 1.4×10^7 m/s. (The charge on an electron is 1.6×10^{-19} C and the magnitude of the Earth's magnetic field is 5×10^{-5} T.)
3. Isotopes of iron (Fe) are to be separated using a mass spectrometer. The applied magnetic field is 45 mT and the applied potential difference is 600 V. The mass of a proton or neutron is 1.66×10^{-27} kg and the charge on a proton is 1.6×10^{-19} C. Find the radii of the paths of ^{54}Fe, ^{56}Fe and ^{57}Fe.

© MoE, RV
Not to be republished

6.4 Magnetic force on current-carrying conductors (long, straight, circular loop)

By the end of this section you should be able to:

- Derive the expression $F = I(l \times B)$.
- Use the expression for the force on a current-carrying conductor in a magnetic field.
- Determine the magnitude and direction of torque acting on a current loop.
- Define magnetic dipole moment.
- Describe the working mechanism of a direct motor.
- Describe and illustrate the magnetic field produced by an electric current in a long straight conductor and in a solenoid.
- Calculate the magnetic field strength of a straight current carrying wire.
- Analyse and predict using the right hand rule the direction of the magnetic field produced when electric current flows through a long straight conductor and a solenoid
- State the Biot–Savart law.
- Apply and use the Biot–Savart law to determine the expression for magnetic field strength of a current element.

The force on a current-carrying conductor in a magnetic field

We know that the force on a charge travelling in a magnetic field is given by

$F = Bqv\sin\theta$.

For a charge travelling a length l in the wire as shown in Figure 6.11, we can substitute $\frac{l}{t}$ for v so the equation becomes $F = Bq\frac{l}{t}\sin\theta$

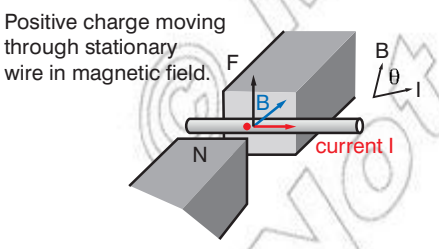


Figure 6.11

We also know that current, I, is $\frac{q}{t}$

So the equation becomes

$F = BIl\sin\theta$

© Not to be published

The direction of the force is perpendicular to both the wire and the magnetic field and is given by the right hand rule as shown in Figure 6.12.

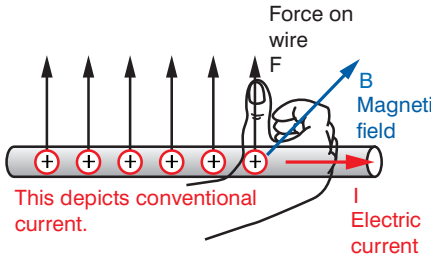


***Figure 6.12** The direction is perpendicular to both wire and magnetic field.*

We can show that the force is a vector (has both magnitude and direction) if we write the equation as follows $\vec{F} = \vec{I}(l \times B)$.

Activity 6.5: The variation of the magnetic field due to a current-carrying conductor

In a small group, investigate the variation of magnetic field due to a current-carrying conductor. Use a ring stand and a clamp to hold a piece of cardboard horizontally. Thread connecting wire through a hole in the cardboard, then connect the wire to a battery, a variable resistor (so that you can vary the current later) and a switch. Place several compasses on the cardboard around the wire.

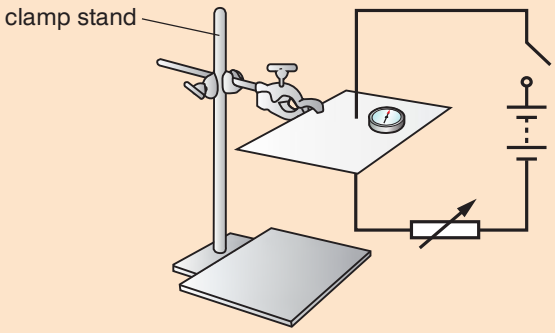


Figure 6.13

Make sure the wire is vertical and close the switch. Observe the behavior of the compasses and sketch the magnetic field lines near the current-carrying wire. Discuss:

- how you decided to draw the lines as you did
- how the field lines that you drew differ from those around a bar magnet.

Repeat this activity with a different current through the wire. How does the field pattern change?

© MoE Ethiopia Not to be sold

Activity 6.6: The effect of the transverse force

Set up the apparatus as shown in Figure 6.14.

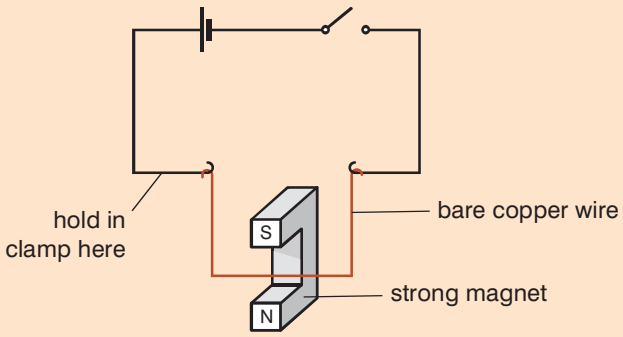


Figure 6.14

Allow a current of up to 5 A to flow through the wire. Observe what happens. Reverse the direction of the current. What happens now?

Worked example 6.6

A student sets up a jumping wire demonstration to impress her younger cousin. She uses a wire with a current of 1.5 A running through it, and a pair of magnets, which have a magnetic field of 0.75 mT. She is a bit careless in setting up and 5 cm of the wire hangs across the field at an angle of 75°.

a) How much force does the wire experience?

b) If it has a mass of 7.5 g, how fast will it accelerate initially?

a)

F (N)	I (A)	l (m)	B (T)	sin θ
?	1.5	0.05	0.75×10^{-3}	0.966

Use $F = BIl\sin\theta$

$= 0.75 \times 10^{-3} \times 1.5 \times 0.05 \times 0.966$

$= 5.44 \times 10^{-5}$ N

b)

F (N)	m (kg)	a (m/s^2)
5.44×10^{-5} N	0.0075	?

Use $F = ma$

$$a = \frac{F}{m}$$

$$= \frac{5.44 \times 10^{-5}}{0.0075}$$

$= 7.25 \times 10^{-3}$ m/s^2

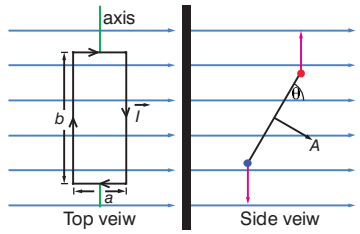


Figure 6.15

The magnitude and direction of torque acting on a current loop

Consider a rectangular loop of wire in a uniform magnetic field, as shown in Figure 6.15.

The sides with length a are parallel to the magnetic field and so they do not experience a force since $\sin\theta = 0$.

The sides with length b are perpendicular to the magnetic field and so each will experience a force of magnitude

$F = BIb$ ($\sin\theta = 1$ for these sides)

However, since the direction of the current is opposite on each side, the direction of the force will be opposite so there is no net force. However, there will be a resultant torque, τ. This torque acts about the axis and is given by

$$\tau = F\frac{a}{2} + F\frac{a}{2} = Fa = IbBa$$

The area of the loop, A, is ab, and so the torque is given by

$$\tau = IBA$$

This is the maximum torque, when the field is in the plane of the loop. In general the torque is given by

$$\tau = I(\vec{A}\times\vec{B})$$

which is also written as

$$\tau = IBA\sin\theta$$

where θ is the angle between the plane of the loop and the field (shown in the side view in Figure 6.15).

You can find the directions of the vectors in this equation using a right hand rule: curl the fingers of your right hand in the direction of the current and your thumb, stuck out, points in the direction of the area vector.

Worked example 6.7

Find the torque on a loop of wire of area 10 cm^2 at 60° to a magnetic field of strength 30 mT with a current of 2 A flowing through it.

τ (N m)	I (A)	B (T)	A (m^2)	$\sin\theta$
?	2	30×10^{-3}	10×10^{-4}	0.866

Use $\tau = IBA\sin\theta$

$= 2 \times 30 \times 10^{-3} \times 10 \times 10^{-4} \times 0.866$

$= 5.196 \times 10^{-5}$ N m

© MoE-TDR Ethiopia Not to be republished

Magnetic dipole moment

A current loop creates a magnetic dipole moment. A dipole moment is defined as

current (I) × area (A)

If we consider a coil of wire consisting of N loops, then the magnetic moment, μ, of such a coil is given by

$\mu = NIA$

The direction of the magnetic moment is given by the right hand rule.

When a magnetic dipole moment is placed in a magnetic field (B), it experiences a torque. This torque is given by the equation

$\tau = \mu \times B$

This can be written as

$\tau = \mu B \sin \theta$

Worked example 6.8

a) Find the magnetic dipole moment on a coil of wire with 100 turns each of area 10 cm^2 at 60° to a magnetic field of strength 30 mT with a current of 2 A flowing through it.

b) Find the torque on the coil described in part (a).

a)

N	I (A)	A (m^2)	μ (A m^2)
100	2	10×10^{-4}	?

Use $\mu = NIA$

$= 100 \times 2 \times 10 \times 10^{-4}$

$= 0.2$ A m^2

b)

τ (N m)	μ (A m^2)	B (T)	sin θ
?	0.2	30×10^{-3}	0.866

Use $\tau = \mu B \sin \theta$

$= 0.2 \times 30 \times 10^{-3} \times 0.866$

$= 5.196 \times 10^{-3}$ N m

The working mechanism of a direct current motor

The principle of the direct current motor is the forced movement of a current carrying loop in a magnetic field. The torque on a current carrying loop will cause it to rotate. If it is free to move then the loop will rotate continuously. You can find the direction of the force using a right hand rule again: the thumb points in the direction of conventional current, the index finger in the direction of the magnetic field and the middle finger in the direction of the force. A direct current motor will have a coil with many turns of wire but Figure 6.16 simplifies the situation and just shows a single rectangular loop.

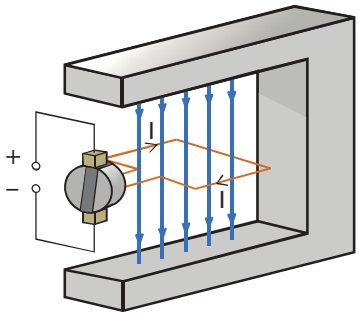


Figure 6.16 *A direct current motor*

Ethiopia published

DID YOU KNOW?

In 2003, the Zettl Lab at the University of Berkeley in California produced a motor which is was less than 500 nm across.

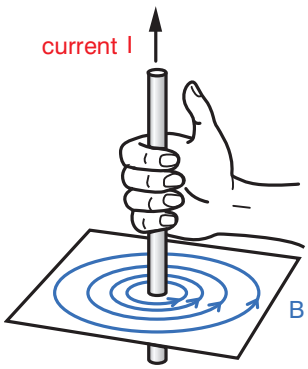


***Figure 6.17** Magnetic field produced by current in a straight conductor*

The magnetic field produced by an electric current in along straight conductor

In Activity 6.5 on page 248, you should have found that the magnetic field around a long straight wire takes the form of concentric circles around the wire. You find the direction of the magnetic field by wrapping the fingers of your right hand around the wire with your thumb in the direction of the current, as shown in Figure 6.17.

The strength of the magnetic field, B, depends on

- the current, I, flowing through the conductor
- the inverse of the distance from the conductor, r (in other words, as the distance from the conductor increases, the strength of the field decreases).

Mathematically, we can write this as

$$B = \frac{kI}{r}$$

where k is a constant term.

It has been found that the value of k depends on the value known as the permeability of free space, μ_0, and the inverse of 2π. So

$$k = \frac{\mu_0}{2\pi}$$

We can see that, for a straight current-carrying conductor

$$B = \frac{\mu_0 I}{2\pi r}$$

Worked example 6.9

In Activity 6.5, you varied the current through your current-carrying wire and plotted the magnetic field. Calculate what the strength of the magnetic field would have been 10 cm from the wire in your experiment if your wire had been carrying a current of 2 A. The permeability of free space is $4\pi \times 10^{-7}$ T m/A

B (T)	μ_0 (T m/A)	I (A)	r (m)
?	$4\pi \times 10^{-7}$	2	0.1

Use

$$B = \frac{\mu_0 I}{2\pi r}$$

$$= \frac{4\pi \times 10^{-7} \times 2}{2\pi \times 0.1}$$

$$= 4 \times 10^{-6} \text{ T}$$

© MoE-FDR Ethiopia Not to be republished

The magnetic force between two wires

If two identical parallel wires each carry current, as shown in Figure 6.18, then each will exert a force F on the other.

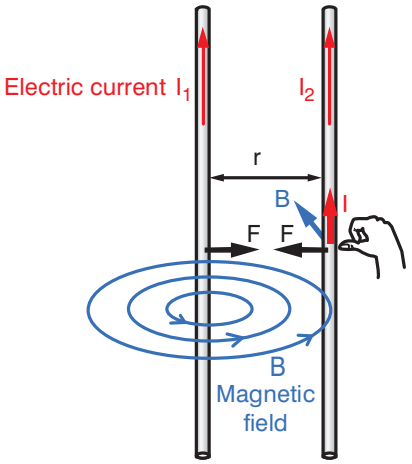


Figure 6.18 *Magnetic force between two wires*

The magnetic field in wire 2 from I_1 is given by

$$B = \frac{\mu_0 I_1}{2\pi r}$$

The force on length Δl of wire 2 is given by

$$F = I_2 \Delta l B$$

The force per unit length in terms of the currents is therefore

$$\frac{F}{\Delta l} = \frac{I_2 \mu_0 I_1}{2\pi r}$$

We can rearrange this to give

$$F = \frac{\mu_0 I_1 I_2 l}{2\pi r}$$

Biot–Savart law and determining the expression for magnetic field strength of a current element

The Biot–Savart law relates magnetic fields to the currents which are their sources (in the same way as Coulomb's law relates electric fields to the point charges which are their sources).

Consider a current-carrying conductor as shown in Figure 6.19.

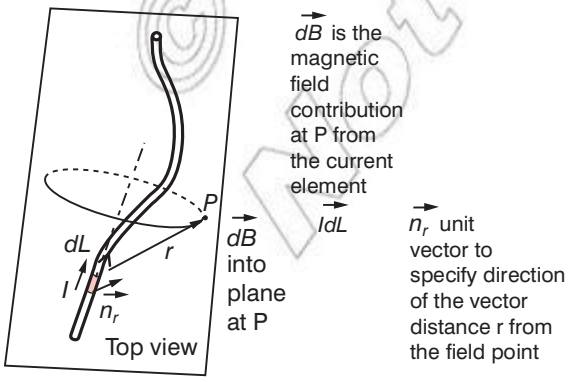


Figure 6.19

© MoE FDR Ethiopia Not to be republished

Each infinitesimal current element I (shown shaded in Figure 6.19) makes a contribution $d\vec{B}$ to the magnetic field at point P. P is perpendicular to the current element and perpendicular to the radius vector r from the current element.

$$d\vec{B} = \frac{\mu_0 I d\vec{l} \times \vec{n}_r}{4\pi r^2}$$

Summary

In this section you have learnt that:

- The expression $F = I(l \times B)$ is used to find the force on a current–carrying conductor in a magnetic field.
- The magnitude of a torque acting on a current loop is given by

  $\tau = IAB\sin\theta$

  where I is current through the loop, A is area of loop, B is magnetic field strength and θ is angle between magnetic field and loop
- A current loop creates a magnetic dipole moment. A dipole moment is defined as:
  current (I) × area (A)
  For a coil of wire consisting of N loops has a magnetic moment, μ, given by $\mu = NIA$
- The direction of the magnetic moment is given by the right hand rule.
- When a magnetic dipole moment is placed in a magnetic field (B), it experiences a torque. This torque is given by the equation:
  $\tau = \mu \times B$
  This can be written as:
  $\tau = \mu B\sin\theta$
- The principle of the direct current motor is the forced movement of a current carrying loop in a magnetic field. The torque on a current carrying loop will cause it to rotate. If it is free to move then the loop will rotate continuously. You can find the direction of the force using a right hand rule again: the thumb points in the direction of conventional current, the index finger in the direction of the magnetic field and the middle finger in the direction of the force.
- The magnetic field produced by an electric current in along straight conductor is concentric circles. It can be calculated using the equation $B = \frac{\mu_0 I}{2\pi r}$.

© MoE - Not for sale

- Its direction is given using the right hand rule.
- The Biot–Savart law can be used to determine the expression for magnetic field strength of a current element $d\vec{B} = \frac{\mu_0 I d\vec{l} \times \vec{n}_r}{4\pi r^2}$.

Review questions

1. a) How much force does a wire with a current of 1.25 A running through it when it is set in magnetic field of 0.5 mT if 7.5 cm of the wire hangs across the field at an angle of 50°?
   b) If the wire has a mass of 10 g, how fast will it accelerate initially?
2. Find the torque on a loop of wire of area 8 cm^2 at 45° to a magnetic field of strength 60 mT with a current of 4.5 A flowing through it.
3. a) Find the magnetic dipole moment on a coil of wire with 500 turns each of area 5 cm^2 at 75° to a magnetic field of strength 25 mT with a current of 1.5 A flowing through it.
   b) Find the torque on the coil described in part a).
4. Describe the principle of the direct current motor.
5. Calculate the strength of the magnetic field 15 cm from a straight current-carrying wire if the wire carries a current of 1 A. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
6. Calculate the force between two parallel wires each of 1 m in length that are 1 m apart when each carries a current of 1 A. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
7. Find the contribution to the magnetic field at a point that is a perpendicular distance 10 cm from a current element of length 10 cm, where the current through the element is 1.5 A. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

© MoE FDRE Ethiopia Not to be republished

6.5 Ampere's law and its application

By the end of this section you should be able to:

- State Ampere's law and use it in solving problems.
- Describe and illustrate the magnetic field produced in a solenoid and predict its direction using the right hand rule.

DID YOU KNOW?

Origin of Ampere's law

In 1820, Oersted discovered that electric currents can induce magnetic fields. Several weeks later Ampere went to a talk in Paris where Oersted's discovery was reported. Ampere began to do detailed experiments to investigate the nature of these induced magnetic fields and their relationship to the electric currents. His research is summed up in Ampere's law.

Ampere's law

Ampere's law states that in any closed loop path, the sum of the length elements times the magnetic field in the direction of the length element is equal to the permeability of free space times the electric current enclosed in the loop, as shown in Figure 6.20.

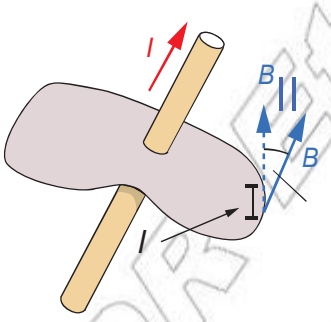


Figure 6.20

Mathematically, we write this as

$$\sum B_{||}\Delta l = \mu_0 I$$

If the angle between the parallel component and the path is θ then the expression becomes

$$\sum B\Delta l\cos\theta = \mu_0 I$$

Ampere's law and the magnetic field inside a wire

We have already found the expression for the magnetic field outside a current-carrying conductor. We can confirm this expression using Ampere's law since, in this case, $\Delta l = 2\pi r$ (the expression for the circumference of a circle). Hence, using

$$\sum B_{||}\Delta l = \mu_0 I$$

and substituting $\Delta l = 2\pi r$, we get

$$B = \frac{\mu_0 I}{2\pi r}$$

We can use Ampere's law to derive an expression for the magnetic field inside a current-carrying conductor. Consider the conductor shown in Figure 6.21.

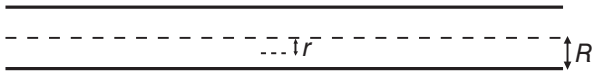


Figure 6.21

© MoE-PDRTI not to be republished

At a distance r from the centre of a conductor of radius R, the current enclosed is given by

$$\frac{Ir^2}{R^2}$$

Using Ampere's law we get

$$\sum B \times 2\pi r = \mu_0 \frac{Ir^2}{R^2}$$

which simplifies to

$$B = \frac{\mu_0 Ir}{2\pi R^2}$$

Worked example 6.10

Find the magnetic field strength inside a conductor of radius 10 mm at a distance of 3 mm from the centre of the conductor when a current of 1.5 A is flowing through the conductor. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

B (T)	μ_0 (T m/A)	I (A)	r (m)	R (m)
?	$4\pi \times 10^{-7}$	1.5	3×10^{-3}	10×10^{-3}

Use

$$B = \frac{\mu_0 Ir}{2\pi R^2}$$

$$= \frac{4\pi \times 10^{-7} \times 1.5 \times 3 \times 10^{-3}}{2\pi \times 10 \times 10^{-3} \times 10 \times 10^{-3}}$$

$$= \frac{9 \times 10^{-10}}{1 \times 10^{-4}}$$

$$= 9 \times 10^{-6} \text{ T}$$

Ampere's law and the magnetic field of a solenoid

A long straight coil of wire, called a solenoid, can be used to generate a magnetic field that is similar to that of a bar magnet. Such coils have many practical applications. The field can be strengthened by adding an iron core. Such cores are typically used in electromagnets.

You can use Ampere's law to find the magnetic field B for a solenoid. Consider the solenoid shown in Figure 6.22.

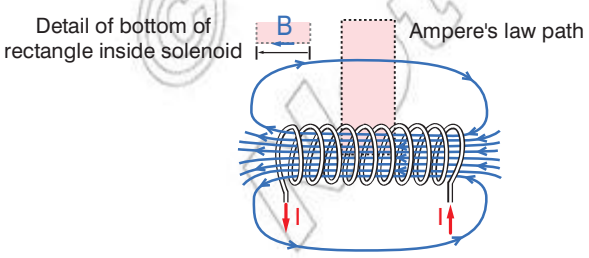


Figure 6.22

If we take a rectangular path so that the length of the side parallel to the solenoid field is length l (shown shaded in Figure 6.22), the contribution to the field from this path is Bl inside the coil where

B is the magnetic field strength. The field can be considered to be perpendicular to the sides of the path so these give negligible contribution.

Using Ampere's law we get, for a solenoid of N turns,

$$Bl = \mu_0 NI$$

$$B = \frac{\mu_0 NI}{l}$$

$$B = \mu_0 nI$$

where n is number of turns per unit length.

We can find the direction of the magnetic field in a solenoid using the right hand rule, as for a current-carrying wire.

Activity 6.7: Making a solenoid

Wrap a 20 cm piece of insulated wire around a pencil several times. Remove the pencil and place the coil of wire on a piece of cardboard. Connect the coil to a battery through two holes in the cardboard. Place several compasses around the coil. Sketch the coil and the compasses in your notebook, then close the switch. Observe the orientation of the compass needles.

Change the shape of the wire and repeat. Compare the shape of the magnetic field for different shapes of wire.

Worked example 6.11

Find the magnetic field inside a solenoid of 1000 turns per unit length with a current of 3 A flowing through it. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

B (T)	μ_0 (T m/A)	n	I (A)
?	$4\pi \times 10^{-7}$	1000	3

Use $B = \mu_0 nI$

$= 4\pi \times 10^{-7} \times 1000 \times 3$

$= 3.768 \times 10^{-3}$ T

Activity 6.8: Investigating the force of attraction between a solenoid and a bar magnet for different values of current through the solenoid

Work in a small group to design and carry out an investigation into the force of attraction between a bar magnet and a solenoid with varying current through the solenoid. What apparatus will you need? What measurements will you need to take? How will you display your results? Write a report on this investigation. Remember that your report should be sufficiently detailed to enable the reader to repeat your procedure and check your results!

© MoE - Not to be republished

Ampere's law and the magnetic field of a toriod

Consider a toroid, as shown in Figure 6.23.

All the loops that make up the toroid contribute magnetic field in the same direction inside the toroid. The direction of the magnetic field can be found using the right hand rule (compare with a solenoid). The current enclosed by the dashed line is just

NI

where N is the number of loops and I is the current in each loop.

Ampere's law then gives the magnetic field since

$B \times 2\pi r = \mu_0 NI$

So

$$B = \frac{\mu_0 NI}{2\pi r}$$

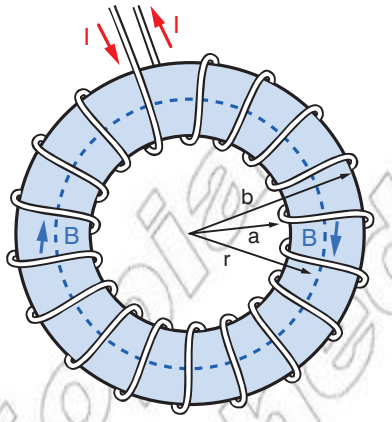


Figure 6.23 A toroid

Worked example 6.12

Find the magnetic field for a toroid of radius 5 cm with 1000 turns per unit length with a current of 3 A flowing through it. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

B (T)	μ_0 (T m/A)	n	I (A)	r (m)
?	$4\pi \times 10^{-7}$	1000	3	0.05

Use

$$B = \frac{\mu_0 NI}{2\pi r}$$

$$= \frac{4\pi \times 10^{-7} \times 1000 \times 3}{2\pi \times 0.05}$$

$$= 0.012 \text{ T}$$

Summary

In this section you have learnt that:

- Ampere's law states that in any closed loop path, the sum of the length elements times the magnetic field in the direction of the length element is equal to the permeability of free space times the electric current enclosed in the loop, as shown in Figure 6.24.

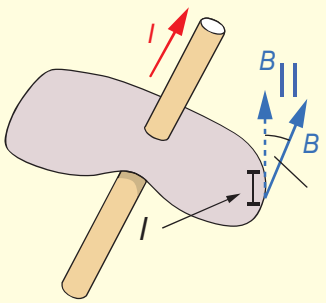


Figure 6.24

Mathematically, we write this as

$$\sum B_{||} \Delta l = \mu_0 I$$

If the angle between the parallel component and the path is θ then the expression becomes

$$\sum B \Delta l \cos\theta = \mu_0 I$$

- Ampere's law can be used to derive the magnetic field strength for a solenoid $B = \mu_0 nI$ where n is number of turns per unit length, the magnetic field inside a conductor

$$B = \frac{\mu_0 Ir}{2\pi R^2}$$

and the magnetic field from a toroid

$$B = \frac{\mu_0 NI}{2\pi r}$$

- The magnetic field produced in a solenoid is similar to that of a bar magnet and its direction can be predicted using the right hand rule.

Review questions

1. State Ampere's law.
2. Find the magnetic field strength inside a conductor of radius 20 mm at a distance of 5 mm from the centre of the conductor when a current of 3 A is flowing through the conductor. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
3. Find the magnetic field inside a solenoid of 500 turns per unit length with a current of 1.5 A flowing through it. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
4. Find the magnetic field for a toroid of radius 3 cm with 500 turns per unit length with a current of 2 A flowing through it. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

6.6 Earth's magnetism

By the end of this section you should be able to:

- Determine the horizontal component of the Earth's magnetic field at a location.
- Resolve the horizontal and vertical components of the Earth's magnetic field.
- Describe how a tangent galvanometer works.

© MoE: Not to be republished

Horizontal and vertical components of the Earth's magnetic field

Except near the equator, the field lines of the Earth's magnetic field are at an angle to the Earth's surface. At the magnetic poles, the field lines pass through the Earth's surface vertically. However, at any other point on the Earth's surface the Earth's magnetic field has a vertical and a horizontal component.

The Earth's magnetic field is a vector quantity; at each point in space it has a strength and a direction.

The strength of the field at the Earth's surface ranges from less than 30 000 nT America and South Africa to over 60 000 nT around the magnetic poles in northern Canada and south of Australia, and in part of Siberia. Near the poles, the field strength diminishes with the inverse square of the distance, i.e. at a distance of R Earth radii it only amounts to $1/R^2$ of the surface field in the same direction, whereas at greater distances, such as in outer space, it diminishes with the cube of the distance. Where the prime meridian intersects with the equator, the field strength is about 31 μT.

DID YOU KNOW?

Generally, the Earth's magnetic field strength is equivalent to 1/30 000th of a tesla. Still, this is enough for birds to navigate by and to keep a compass hand pointed north. The magnetic field of Jupiter, the largest planet in the solar system, is about ten times stronger than Earth's, or 1/3000th of a tesla.

The tangent galvanometer

The tangent galvanometer (TG) is an instrument for measuring the strength of an electrical current in terms of the magnetic field it produces. A TG consists of a circular coil of insulated copper wire wound on a circular non magnetic frame. The frame is mounted vertically on a horizontal base provided with levelling screws on the base. The coil can be rotated on a vertical axis passing through its centre. A compass box is mounted horizontally at the centre of a circular scale. The compass box is circular in shape. It consists of a tiny, powerful magnetic needle pivoted at the centre of the coil. The magnetic needle is free to rotate in the horizontal plane. The circular scale is divided into four quadrants. Each quadrant is graduated from 0° to 90°. A long thin aluminium pointer is attached to the needle at its centre and at right angle to it. To avoid errors due to parallax a plane mirror is mounted below the compass needle.



Figure 6.25 *A tangent galvanometer*

Current flowing through a coil of wire generates a magnetic field at the centre of a coil, and this field deflects a magnetic compass needle. The instrument derives its name from the fact that the current is proportional to the tangent of the angle of the needle's deflection. When current is passed through the TG a magnetic field is created at its centre. This field is given by

$$B_c = \frac{\mu_0 NI}{2R}$$

where N is the number of turns of wire in the coil, I is the current through it and R is the radius of the coil.

© Not to be republished

DID YOU KNOW?

The Magnetic Observatory of Addis Ababa has been operating since January 1958. In August 1997, the Institut de Physique du Globe de Paris (IPGP) installed a new magnetic station in Addis Ababa as part of its network 'Observatoire Magnetique Planetaire'. In August 2004, a new vector magnetometer VM391 was installed. Its purpose is to provide ground-based, calibrated values of the Earth's magnetic field at a specific location for scientific research and practical applications.

If the TG is set such that the plane of the coil is along the magnetic meridian, i.e. B_c is perpendicular to the horizontal component of the Earth's magnetic field, the needle rests along the resultant, as shown in Figure 6.26.

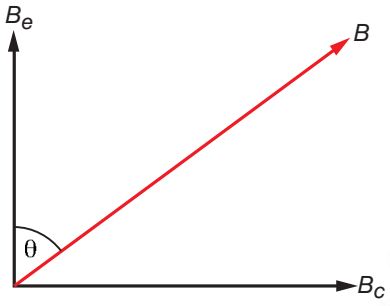


Figure 6.26

Because a compass aligns itself with the lines of force of the magnetic field within which it is placed, a compass can be used to find the angle θ between B_e and B. If the compass is first aligned with the magnetic field of the Earth and current is supplied to the coils, then the compass needle will undergo an angular deflection aligning itself with the vector sum of the Earth's field and the field due to the coils. This angular deflection is θ.

The horizontal component of the Earth's magnetic field can be expressed as

$$B_e = \frac{B_c}{\tan\theta}$$

Activity 6.9: Using a tangent galvanometer to determine the strength of the Earth's magnetic field at your location

If possible, use a tangent galvanometer to determine the strength of the Earth's magnetic field at your location. If this is not possible, carry out some research into the Magnetic Observatory at Addis Ababa.

© MoE-Ethiopia Not to be republished

Summary

In this section you have learnt that:

- The Earth's magnetic field is a vector quantity; at each point in space it has a strength and a direction.
- Except near the equator, the field lines of the Earth's magnetic field are at an angle to the Earth's surface. At the magnetic poles, the field lines pass through the Earth's surface vertically. However, at any other point on the Earth's surface the Earth's magnetic field has a vertical and a horizontal component.
- The strength of the field at the Earth's surface ranges from less than 30 000 nT in America and South Africa to over 60 000 nT around the magnetic poles in northern Canada and south of Australia, and in parts of Siberia.
- The tangent galvanometer (TG) is an instrument for measuring the strength of an electrical current in terms of the magnetic field it produces.
- The current is proportional to the tangent of the angle of the needle's deflection. When current is passed through the TG a magnetic field is created at its centre. This field is given by

  $$B_c = \frac{\mu_0 NI}{2R}$$

  where N is the number of turns of wire in the coil, I is the current through it and R is the radius of the coil.
- If the TG is set such that the plane of the coil is along the magnetic meridian, i.e. B_c is perpendicular to the horizontal component of the Earth's magnetic field, the needle rests along the resultant, as shown in Figure 6.27.

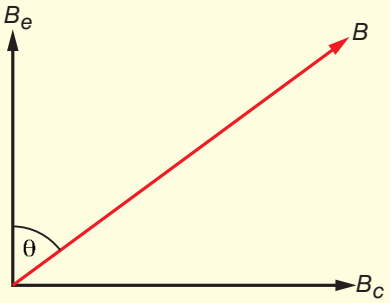


Figure 6.27

- The horizontal component of the Earth's magnetic field can be expressed as

  $$B_e = \frac{B_c}{\tan \theta}$$

Namibia Open University Published

Review questions

1. What instrument could be used to determine the horizontal component of the Earth's magnetic field?
2. At a particular location, $\theta = 70°$. If coils of radii 10 cm with 1000 turns and a current of 5 A were used to determine the horizontal component of the Earth's magnetic field at this location, what measurement would be given for the strength of this component at this location? The permeability of free space is $4\pi \times 10^{-7}$ T m/A.

End of unit questions

1. Describe an experiment to demonstrate the three-dimensional nature of the magnetic field around a bar magnet.
2. A bar magnet causes a magnetic field with a strength of 30 mT at an angle of 45° to a region of area 15 cm^2. How much flux will be contained by this region?
3. What does the term 'magnetism' describe?
4. Explain the difference between paramagnetic and ferromagnetic materials.
5. Outline the dynamo theory which can be used to explain the Earth's magnetic field.
6. Find the speed of an electron travelling at an angle of 40° to the Earth's magnetic field that experiences a force of 8×10^{-17} N. (The charge on an electron is 1.6×10^{-19} C and the magnitude of the Earth's magnetic field is 5×10^{-5} T.)
7. Outline J.J. Thompson's experiment of charge to mass ratio.
8. Outline the principles of a mass specrometer.
9. Isotopes of carbon (C) are to be separated using a mass spectrometer. The applied magnetic field is 45 mT and the applied potential difference is 600 V. The mass of a proton or neutron is 1.66×10^{-27} kg and the charge on a proton is 1.6×10^{-19} C. Find the radii of the paths of ^{12}C, ^{13}C and ^{14}C.
10. Describe how you would investigate the variation of the magnetic field due to a current-carrying conductor.
11. a) A straight piece of conducting wire, 20 cm long, lies at 70° a magnetic field of 1.8×10^{-5} T. A current, *I*, is allowed to flow through it and it experiences a force of 0.01 N. Calculate the value of *I*.

    b) The wire in part (a) is bent into a rectangular shape so that it has an area of 24 cm^2. Find the torque on the loop of wire 70° to the magnetic field of strength 1.8×10^{-5} T when the same current as in part (a) flows through it.

© MoE: Ethiopia Not to be republished

c) Find the magnetic dipole moment on a coil of wire with 1000 turns each of area 24 cm^2 at 70° to a magnetic field of strength 1.8×10^{-5} T with the same current as part (a) flowing through the coil.

12. Describe the principle of the direct current motor
13. How far from a straight current–carrying wire carrying a current of 1 A is the strength of the magnetic field 1.5×10^{-6} T? The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
14. Show how to derive the expression for the magnetic field between two wires.
15. Find the contribution to the magnetic field at a point that is a perpendicular distance 15 cm from a current element of length 20 cm, where the current through the element is 5 A. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
16. State Ampere's law.
17. How far from the centre of a conductor of radius 10 mm with a current of 1.5 A flowing through it is the magnetic field strength 5×10^{-6} T? The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
18. Consider the solenoid shown in Figure 6.28

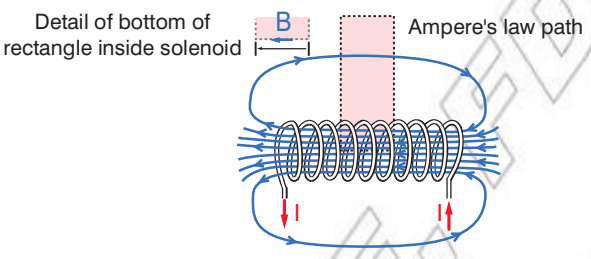


Figure 6.28

Derive the expression for the magnetic field of this solenoid.

19. What current is needed through a solenoid of 350 turns if the magnetic field inside it is to be 9×10^{-4}? The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
20. Find the magnetic field for a toroid of radius 3 cm with 200 turns per unit length with a current of 2 A flowing through it. The permeability of free space is $4\pi \times 10^{-7}$ T m/A.
21. What is the range of the strength of the magnetic field at the earth's surface.
22. Describe how a tangent galvanometer works.

© MoE-EDRI Ethiopia Not to be republished

Electromagnetic induction and a.c. circuits

Unit 7

Contents

Section	Learning competencies
7.1 Phenomena of electromagnetic induction (page 268)	• Define magnetic flux. • Use $\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA\cos\theta$ to solve related problems. • use the terms induced emf, back e.m.f, magnetic flux, flux linkage, eddy current • Describe experiments to investigate the factors that determine the direction and magnitude of an induced e.m.f. • Use an expression for the induced e.m.f. in a conductor moving through a uniform magnetic field by considering the forces on the charges • State the laws of electromagnetic induction • Use the laws of electromagnetic induction which predict the magnitude and direction of the induced e.m.f. • Use $\varepsilon = -N\frac{\Delta\emptyset}{\Delta t}$ to solve related problems. • Solve problems involving calculations of the induced emf, the induced current. • Analyse and describe electromagnetic induction in qualitative terms. • Apply Lenz's law to explain, predict and illustrate the direction of the electric current induced by a changing magnetic field, using the right-hand rule. • Describe the effects of eddy currents in large pieces of conducting materials. • Define the terms self-inductance, L, mutual inductance, M, and henry, H. • State the factors that determine the magnitude of self-inductance and mutual inductance. • Derive an expression for the inductance of a solenoid ($L = n^2 s\mu_0 A$). • Derive and use the expression for the energy stored in an inductor ($PE_B = \frac{1}{2}LI^2$). • Define magnetic energy density.
7.2 Alternating current (a.c.) generator and transformers (page 281)	• Compare direct current (d.c.) and alternating current (a.c.) in qualitative terms. • Derive the expression for the emf induced in a rotating coil $\varepsilon = \omega NBA\sin\omega t$. • Draw a schematic diagram for a simple a.c. generator. • Explain the working mechanism of a generator. • Draw a schematic diagram of a transformer. • Derive the transformer equation $\frac{V_1}{V_2} = \frac{N_1}{N_2} = \frac{I_2}{I_1}$ from Faraday's law. • Explain the importance of alternating current in the transmission of electrical energy.

Contents

Section	Learning competencies
7.3 Alternating current (a.c.) (page 287)	• Explain what is meant by r.m.s. values. • Apply the relationship between r.m.s. and peak values for the current and potential difference for a sinusoidal waveform. • Identify that the current and voltage are in phase in a resistor in an a.c. circuit. • Explain the behaviour of a capacitor in an a.c. circuit. • Derive the expression for the instantaneous current and voltage in a resistive and capacitive circuit. • Identify that the current leads the voltage by $\frac{\pi}{2}$ in a capacitor in an a.c. circuit. • Draw phasor diagrams for resistive and capacitive circuits. • Define capacitive reactance. • Use the terms: r.m.s. current, r.m.s. potential difference, peak current, peak potential difference, half cycle average current, phase difference, phase lag, phase lead. • Use the terms: reactance, impedance, power factor with their correct scientific meaning. • Define the power factor in an a.c. circuit. • Identify that the voltage leads the current by $\frac{\pi}{2}$ in an inductive circuit. • Explain the behaviour of an inductor in an a.c. circuit. • Derive the expression for the instantaneous current/voltage in an inductor in an inductive circuit. • Define inductive reactance. • Describe the behaviour of an RL circuit. • Describe the behaviour of an LC circuit. • Describe the behaviour of RLC circuits. • Derive an expression for the impedance of RLC circuits. • Draw phasor diagrams for RLC circuits. • Solve problems involving the magnitude and phase of current and applied p.d. in a.c. circuits which include resistors, capacitors and inductors.
7.4 Power in a.c. circuits (page 304)	• Show that the average power in an a.c. capacitive circuit is zero. • Derive the expression for the average power in an a.c. inductive circuit. • Derive the expression for the average power in an a.c. RLC circuit. • Distinguish between real, apparent and ideal power of an RLC circuit.

No

7.1 Phenomena of electromagnetic induction

By the end of this section you should be able to:

- Define magnetic flux.
- Use $\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA\cos\theta$ to solve related problems.
- Use the terms induced e.m.f, back e.m.f, magnetic flux, flux linkage, eddy current.
- Describe experiments to investigate the factors that determine the direction and magnitude of an induced e.m.f.
- Use an expression for the induced e.m.f. in a conductor moving through a uniform magnetic field by considering the forces on the charges.
- State the laws of electromagnetic induction.
- Use the laws of electromagnetic induction that predict the magnitude and direction of the induced e.m.f.
- Use $\varepsilon = -N\frac{\Delta \emptyset}{\Delta t}$ to solve related problems.
- Solve problems involving calculations of the induced e.m.f, the induced current.
- Analyse and describe electromagnetic induction in qualitative terms.
- Apply Lenz's law to explain, predict and illustrate the direction of the electric current induced by a changing magnetic field, using the right-hand rule.
- Describe the effects of eddy currents in large pieces of conducting materials.
- Define the terms self inductance, L, mutual inductance, M, and henry, H.
- State the factors that determine the magnitude of self inductance and mutual inductance.
- Derive an expression for the inductance of a solenoid ($L = n^2 s \mu_0 A$).
- Derive and use the expression for the energy stored in an inductor ($PE_B = \frac{1}{2}LI^2$).
- Define magnetic energy density.

KEY WORDS

magnetic flux *is represented by magnetic field lines in diagrams that show magnetic fields*

Magnetic flux

In Unit 6, you were introduced to **magnetic flux**. You learnt that magnetic flux is represented by magnetic field lines in diagrams which show magnetic fields. The strength of a magnetic field is represented by the symbol B, and is also called magnetic flux density.

© MoE - Not for sale

In the general case where the area of flux is at an angle θ to the magnetic field, as shown in Figure 7.1, we use the scalar product

$\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA\cos\theta$

to calculate the magnetic flux.

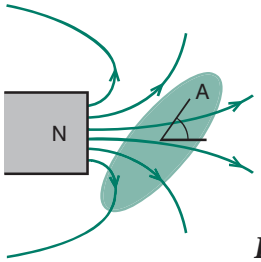


***Figure 7.1** Area of flux*

Worked example 7.1

Calculate the magnetic flux when a magnetic field of strength 5×10^{-5} T passes through an area of 10 cm^2 that is at an angle of 60° to the magnetic field.

Φ_B (Wb)	B (T)	A (m^2)	cos θ
?	5×10^{-5}	10×10^{-4}	0.5

Use $\Phi_B = BA\cos\theta$

$= 5 \times 10^{-5} \times 10 \times 10^{-4} \times 0.5$

$= 2.5 \times 10^{-8}$ Wb

Induced e.m.f.

In Grade 10 you explored how it is possible to induce an e.m.f. in a circuit. This section revises and extends the ideas that you met then.

Activity 7.1: What do you remember about induced e.m.f.s?

In a small group, write down all that you can remember from Grade 10 about induced e.m.f.s. Share your thoughts with the rest of your class.

In Unit 6, you learnt that when a charged particle moves in a magnetic field, it experiences a force. Newton's third law of motion tells us that this force must have an equal and opposite force. This pair of forces occurs whenever there is relative motion between a charge and a magnetic field; the velocity term in the expression $F = Bqv\sin\theta$ that you met in Unit 6 refers to the relative (perpendicular) velocity between the magnetic field and the charge.

This means that, if a magnetic field moves (or changes) near a wire, the electrons in the wire will feel a force which will tend to make the conduction electrons move through the wire. This movement of conduction electrons is an **induced e.m.f.** and, if the wire is part of a complete circuit, then the electrons will move, producing a current.

KEY WORDS

induced e.m.f. *an e.m.f. produced when a magnetic field moves (or changes) near a wire*

Activity 7.2: Demonstrating an induced e.m.f. and investigating factors that influence its magnitude

Work in the same group as you did for Activity 7.1. Set up the apparatus shown in the diagram.

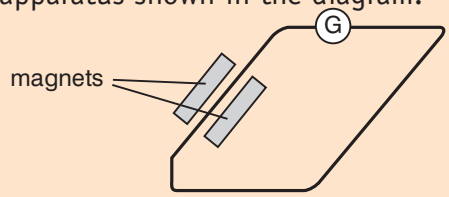


Figure 7.2

Move the wire between the magnets. Observe what happens to the needle on the galvanometer.

Now investigate what happens to the size of the induced e.m.f. when you vary the conditions of the experiment.

- Try changing the strength of the magnetic field. How does this affect the induced e.m.f.?
- Try moving the wire through the magnetic field at different speeds. How does this affect the induced e.m.f.?
- Now make a coil of wire, insert it in the circuit and move the coil through the magnetic field. How does this affect the induced e.m.f.?

From your results in Activity 7.2, you should have realised that the magnitude of the induced e.m.f. depends on:

- the strength of the magnetic field (stronger magnetic field means larger e.m.f.)
- the speed at which you move the wire through the magnetic field (greater speed means greater e.m.f.)
- if you coil the wire so that more wire is influenced by the magnetic field, you will induce a greater e.m.f.

As a straight wire moves in a magnetic field as shown in Figure 7.3 it cuts the flux in the area swept out by the wire.

The area swept out per second by a wire of length l moving at a velocity $v = lv$

The flux cut per second is therefore Blv. This gives the induced e.m.f. as

$\varepsilon = Blv$

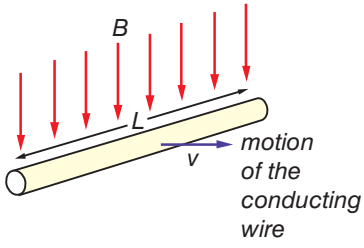


Figure 7.3

© MoE; FET Not to be sold

Flux linkage

KEY WORDS

magnetic flux linkage *the amount of magnetic flux that interacts with a coil of wire*

You know that if a magnetic field can influence a greater length of the wire (such as when the wire is coiled) the induced e.m.f. is greater. The amount of magnetic flux that interacts with a coil of wire is called the **magnetic flux linkage**. Magnetic flux linkage is the product of the number of turns of wire, N, and the flux in that region, Φ, so we can write

flux linkage = $N\Phi$

The units are weber turns.

Since $\Phi = BA\cos\theta$ we can write

flux linkage = $BAN\cos\theta$, where θ is the angle between the area of flux and the magnetic field.

Worked example 7.2

A student takes a wire and coils it into 20 circular coils with a radius of 2.5 cm. He then passes it through a magnetic field of strength 50 mT at an angle of 60°. What is the flux linkage?

Flux linkage (weber turns)	B (T)	A (m^2)	N	cos θ
?	50×10^{-3}	$\pi \times (2.5 \times 10^{-4})^2$	20	0.5

Use flux linkage = $BAN\cos\theta$

$= 50 \times 10^{-3} \times \pi \times (2.5 \times 10^{-4})^2 \times 20 \times 0.5$

$= 50 \times 10^{-3} \times 1.9625 \times 10^{-3} \times 20 \times 0.5$

$= 9.81 \times 10^{-4}$ weber turns

The laws of electromagnetic induction

Your results from Activity 7.2 can be summarised in Faraday's law of electromagnetic induction:

the magnitude of an induced e.m.f., ε, is proportional to the rate of change of flux.

Mathematically, we can write this as

$$\varepsilon = k\frac{\Delta\Phi}{\Delta t}$$

where k is the constant of proportionality.

Activity 7.3: towards Lenz's law

Work in a small group. Take a piece of copper tube and drop a) a magnet and b) a piece of non-magnetic metal of the same size as the magnet through the copper tube. Compare the times that it takes the magnet and the non-magnetic metal to fall through the tube. What do you notice? Try and explain your observations before reading on. (Hint: think of the copper tube as stacked coils of copper wire.)

© Moet to Ethiopia Not to be published

In Activity 7.3, you will have found that the magnet fell through the copper tube more slowly than the non-metallic metal, even though they were the same size and so the friction forces should have been the same on both since copper is non-magnetic.

However, if you imagine that the copper tube is a series of coils of copper wire all stacked on top of each other, you can see that, as the magnet fell through the tube, it induced an e.m.f. in each coil, which caused a small current to flow in the tube. This current then generated an electromagnetic field, which interacted with the falling magnet. The direction of the electromagnetic field determined whether the magnet slowed down or was repelled faster down the tube. If the magnet had been repelled faster down the tube, then its final kinetic energy would have been greater than the gravitational potential energy that it had at the start. From the law of conservation of energy, we know that this is impossible, so the direction of the induced electromagnetic field must have been such as to oppose the motion of the magnet (which of course induced it in the first place)!

This is summarised in Lenz's law:

the direction of the induced e.m.f. is such as to oppose the change creating it.

If we include this law about the direction of the induced e.m.f. with our mathematical expression for Faraday's law we get

$$\varepsilon = -\frac{\Delta\Phi}{\Delta t}$$

For a coil with N turns this can be written as

$$\varepsilon = -\frac{N\Delta\Phi}{\Delta t}$$

and since $\Phi = BA$ for a coil of area A

$$\varepsilon = -\frac{N\Delta(BA)}{\Delta t}$$

The right hand rule for induced e.m.f.s

The direction of an induced e.m.f. can be found using the right hand rule as shown in the diagram.

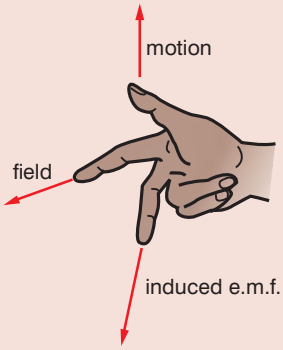


Figure 7.4

Worked example 7.3

The coil of wire in worked example 7.2 on page 271 is moved from the magnetic field to a place completely outside the magnetic field in a time of 0.3 seconds. What e.m.f. is induced in the coil?

ε (V)	$N\Delta\Phi$ (weber turns)	Δt (s)
?	9.81×10^{-4}	0.3

Use

$$\varepsilon = -\frac{N\Delta\Phi}{\Delta t}$$

$$= -\frac{9.81 \times 10^{-4}}{0.3}$$

$$= 3.27 \times 10^{-3}\ \text{V}$$

© Not to be republished

Worked example 7.4

A coil has an area of 16 cm² and has 450 turns. Calculate the induced e.m.f. in the coil when the flux density through the coil changes at a rate of 0.5 T/s.

ε (V)	N	B (T)	A (M²)
?	450	0.5	16×10^{-4}

Use

$$\varepsilon = -\frac{N\Delta\Phi(BA)}{\Delta t}$$

$$= -\frac{450 \times 0.5 \times 16 \times 10^{-4}}{1}$$

$$= 0.36 \text{ V}$$

Mutual inductance

We know that where there is relative motion between a conductor, or coil, and the field of a permanent magnet, e.m.f. is induced. However, the magnetic field that induces the e.m.f. could be produced by another coil as shown in Figure 7.5. This is known as **mutual inductance**. The unit of inductance is the **henry** (H).

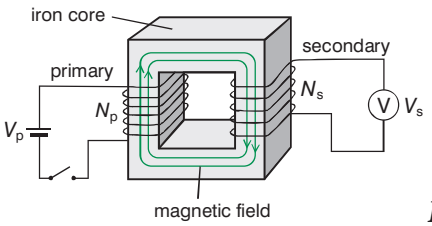


Figure 7.5

In such a set up, virtually all the magnetic field generated by the primary coil would interact with the secondary coil since iron is very good at carrying magnetism. When the switch is closed in the primary circuit, a magnetic field as a result of the primary coil is suddenly produced. This change in magnetic flux linkage will induce an e.m.f. in the secondary coil which may be observed on the voltmeter. Once the magnetic field has settled down in the primary coil and there is no further change, there would be no further induced e.m.f. and the reading on the voltmeter would return to zero. When the switch in the primary circuit is opened again, an e.m.f. is again induced in the secondary coil, but this time it is in the reverse direction.

Eddy currents

We know that a voltage is induced in a coil when there is a change in magnetic flux. If a voltage is induced, there will also be an induced current in the coil. If there is a changing magnetic flux in a solid metallic object, then there will also be an induced voltage and current in the metallic object. The induced current is called an **eddy current**. The currents circulate in complete loops. The charged particles travel in one direction and then the other as the magnetic field changes direction.

KEY WORDS

mutual inductance *when a change in the magnetic field due to one coil induces an e.m.f. in another coil*

henry *the unit of inductance 1 H = 1 Wb/A*

eddy current *current induced in a solid metallic object when there is a change of magnetic flux*

The rate of change of magnetic flux in coil 2, $\emptyset_{21}$, is proportional to the rate of change of current in coil 1 and $N_2 \frac{\Delta\Phi_{21}}{\Delta t} = M_{21} \frac{\Delta I_1}{\Delta t}$

© MoE FDRE Ethiopia Not to be republished

DID YOU KNOW?

Eddy currents are used increasingly in security checks at airports. The detector produces a magnetic field which will induce eddy currents in any metal objects which pass through it, such as keys, coins, etc.

A solid metallic object can be thought of as being built of many rings of different sizes as shown in Figure 7.6.

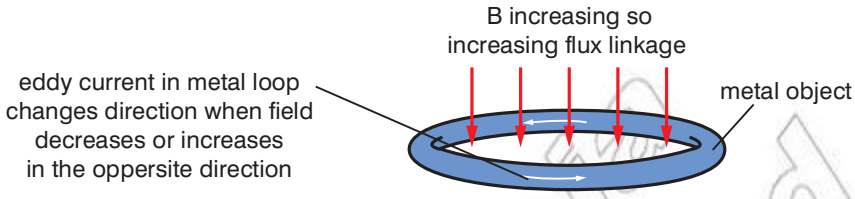


Figure 7.6

The change in flux linkage in each loop, and therefore the induced voltage, is proportional to the area (πr^2). The resistance of the material in each loop is proportional to the length of the loop ($2\pi r$). The induced current is given by

induced voltage/resistance

so the induced current in each loop is proportional to $\frac{\pi r^2}{\pi r} = r$.

Eddy currents produce a heating effect (since $P = I^2R$). This principle is used in induction welding and in many manufacturing processes.

Activity 7.4: Researching applications of eddy currents

In a small group, carry out some research into the applications of eddy currents. Present your findings to your class in a form of your choice.

Self-induction

Any current, I, in an electrical circuit, produces a magnetic field and hence generates a magnetic flux, Φ, acting on the circuit. According to Lenz's law, this magnetic flux tends to act to oppose changes in the flux by generating a voltage (**back e.m.f.**) in the circuit which counters or tends to reduce the rate of change of current. The ratio of the magnetic flux (Φ) times turns of wire (N) to the current is called the **self-inductance** (L) of the circuit. In symbols this is written as

$$L = \frac{\Phi N}{I}$$

The self-inductance is usually referred to as the inductance of the circuit.

KEY WORDS

back e.m.f. *ca voltage induced by a changing magnetic field. It opposes the current (and thus the e.m.f.) that induced it.*

self-inductance *the ratio of the magnetic flux (Φ) times turns of wire (N) to the current*

The factors which determine the magnitude of self-inductance and mutual inductance

From the equation for self-inductance, you can see that if you increase the magnetic flux or the number of turns of wire, you will increase the self-inductance of the circuit. Similarly, if you decrease the current you will increase the self-inductance of the circuit. So the factors which determine the magnitude of self-inductance are

- magnetic flux
- number of turns on coil
- current through the coil (decreasing current increases inductance).

© Not to be republished

Worked example 7.5

Find the self-inductance in a coil where the magnetic field is 30 mT, the area of the coil is 5 cm^2, there are 50 turns on the coil and the current through the coil is 1.5 A.

B (T)	A (m^2)	Φ (BA)	N	I (A)	L (H)
30×10^{-3}	5×10^{-4}	1.5×10^{-5}	50	1.5	?

Use

$$L = -\frac{\Phi N}{I}$$

$$= \frac{1.5 \times 10^{-5} \times 50}{1.5}$$

$$= 5 \times 10^{-4} \text{ H}$$

Inductance and induced e.m.f.

The term inductor is used to describe a circuit element which possesses the property of inductance. A coil of wire is a very common inductor. If we draw a coil of wire as shown in Figure 7.7, we can understand why a voltage is induced in a wire carrying a changing current.

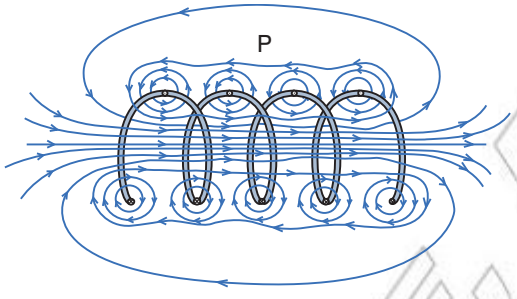


Figure 7.7

The changing current running through the coil creates a varying magnetic field in and around the coil, as shown in Figure 7.7. When the current increases in one loop its magnetic field will expand and cut across some or all of the surrounding loops, inducing a voltage in these loops. Thus, when the current is changing throughout the coil, a voltage is induced throughout the coil.

We can see that we can increase the induced voltage by

- increasing the number of turns in the coil
- increasing the rate of change of magnetic flux.

Faraday's law tells us that

$$\varepsilon = -N\frac{\Delta\Phi}{\Delta t}$$

We also know that

$$L = -\frac{\Phi N}{I}$$

Since the rate of change of magnetic flux is related to the rate of change of current through the coil, (rate of change of flux is related to rate of change of voltage, which is proportional to the current through the coil) and it is easier to measure a rate of change of

© Ministry of Education FDRE Ethiopia Not to be republished

current than a rate of change of magnetic flux, and LI = ΦN, in practice these two equations are combined to give

$$\varepsilon = -L\frac{\Delta I}{\Delta t}$$

Worked example 7.6

Find the induced voltage in an inductor of 40 mH when the current is changing at a rate of 0.5 A/s.

ε (V)	L (H)	$\frac{\Delta I}{\Delta t}$ (A/s)
?	40×10^{-3}	0.5

Use

$$\varepsilon = -L\frac{\Delta I}{\Delta t}$$
$$= 40 \times 10^{-3} \times 0.5$$
$$= -20 \times 10^{-3}\ \text{V}$$

The inductance of a solenoid

Consider a coil of N turns and length l in a circuit as shown in Figure 7.8.

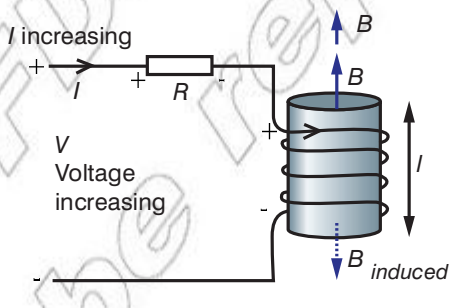


Figure 7.8

For a fixed area, A, and a changing current, I, Faraday's law becomes

$$\varepsilon = -N\frac{\Delta\Phi}{\Delta t} = -NA\frac{\Delta B}{\Delta t}$$

We learnt in Unit 6 that the magnetic field of a solenoid is given by

$$B = \mu\frac{N}{l}I$$

For a long coil the e.m.f. is approximated by

$$\varepsilon = -N\frac{\mu N^2 A\Delta I}{l\Delta t}$$

We also know that

$$\varepsilon = -L\frac{\Delta I}{\Delta t}$$

If we equate these two expressions we get

$$L = \frac{\mu N^2 A}{l}$$

For a coil with n turns per unit length with an air core this expression simplifies to

$$L = \mu_0 n^2 A$$

© MoE, Free State
Not to be reproduced

Worked example 7.7

Find the inductance of an air cored solenoid of 500 turns per unit length and area 5 cm^2. The permeability of free space is $4\pi \times 10^{-7}$ T m^2/A.

L (H)	μ_0 (T m^2/A)	n	A (m^2)
?	$4\pi \times 10^{-7}$	500	5×10^{-4}

Use $L = \mu_0 n^2 A$

$= 4\pi \times 10^{-7} \times 500^2 \times 5 \times 10^{-4}$

$= 4\pi \times 10^{-7} \times 2.5 \times 10^5 \times 5 \times 10^{-4}$

$= 1.57 \times 10^{-4}$ H

The energy stored in an inductor

Consider an inductor in a circuit, as shown in Figure 7.9.

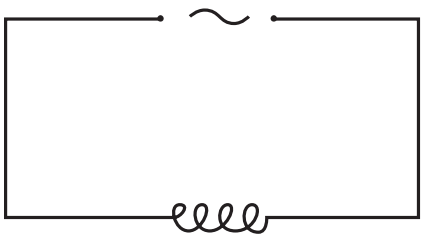


Figure 7.9

When an electric current flows through the inductor, we know that there is an induced voltage given by

$$\varepsilon = -L\frac{\Delta I}{\Delta t}$$

When the current is flowing through the inductor, there is energy stored in the magnetic field, which we give the symbol PE_B (it is potential energy PE stored by a magnetic field, B).

The instantaneous power that must be supplied to the inductor to initiate the current in the conductor, P, is given by

$$P = I\varepsilon = -LI\frac{\Delta I}{\Delta t}$$

We can find the energy stored when there is a final current I_F at time t by integrating the expression for power like this

$$PE_B = \int_0^t P\,dt = \int_0^{I_F} LI\,dI$$

$$= \frac{1}{2}LI^2$$

Think about this...

Integrating $\int_0^{I_F} LI\,dI$

Compare this integral with $\int kx\,dx$.

If you integrate $\int kx\,dx$ the result is $\frac{1}{2}kx^2$.

© MoE FDR Ethiopia Not to be republished

Worked example 7.8

Find the energy stored in the inductor in worked example 7.7 when a current of 2 A flows through it.

PE_B (J)	L (H)	I (A)
?	1.57×10^{-4}	2

Use $PE_B = \frac{1}{2} LI^2$

$= 0.5 \times 1.57 \times 10^{-4} \times 4$

$= 3.14 \times 10^{-4}$ J

Magnetic energy density

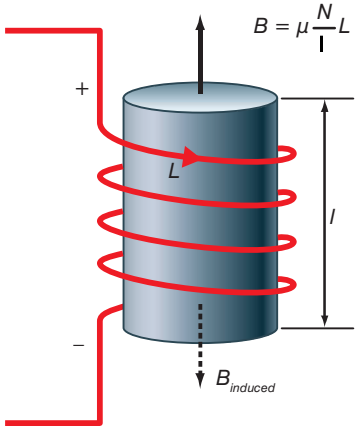


Figure 7.10

Magnetic energy density is defined as

$u_B = \frac{\text{energy}}{\text{volume}}$

Consider an inductor of length l and area of cross section A as shown in Figure 7.10.

We know that the energy stored in the inductor = $PE_B = \frac{1}{2} LI^2$

We also know that $L = \frac{\mu_0 N^2 A}{l}$ and that $I^2 = \frac{B^2 l^2}{\mu_0^2 N^2}$

If we substitute these values into $PE_B = \frac{1}{2} LI^2$ and simplify we get

$PE_B = \frac{AB^2 l}{2\mu_0}$

The volume is Al so the energy density $u_B = \frac{\text{energy}}{\text{volume}}$ is given by $\frac{AB^2 l}{2\mu_0} \times \frac{1}{Al}$

This simplifies to

$u_B = \frac{B^2}{2\mu_0}$

KEY WORDS

magnetic energy density *the energy per unit volume of a magnetic field*

Worked example 7.9

Find the energy density for an inductor with an air core with a magnetic field of 0.5 T. The permeability of free space is $4\pi \times 10^{-7}$ T m²/A.

η_B (J/m³)	B (T)	μ (T m²/A)
?	0.5	$4\pi \times 10^{-7}$

Use $\eta B = \frac{B^2}{2\mu}$

$= \frac{0.5^2}{2 \times 4\pi \times 10^{-7}}$

$= \frac{0.25}{2.512 \times 10^{-6}}$

$= 99\ 522$ J/m³

© MOE FOR Ethiopia Not to be republished

Summary

In this section you have learnt that:

- Magnetic flux is represented by magnetic field lines in diagrams which show magnetic fields.
- The strength of a magnetic field is represented by the symbol B, and is also called magnetic flux density.
- We use the scalar product $\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA\cos\theta$
  Where the area of flux is at an angle θ to the magnetic field to calculate the magnetic flux.
- Induced e.m.f. is an e.m.f. produced when a magnetic field moves (or changes) near a wire.
- Back e.m.f. is an e.m.f. caused by a changing magnetic field. It opposes the current (and thus e.m.f.) that induced it.
- Flux linkage is the amount of magnetic flux that interacts with a coil of wire.
- Flux linkage = $BAN\cos\theta$, where θ is the angle between the area of flux and the magnetic field
- Eddy current is current induced in a solid metallic object when there is a change of magnetic flux.
- The factors that determine the direction and magnitude of an induced e.m.f. are
  - the strength of the magnetic field (stronger magnetic field means larger e.m.f.)
  - the speed at that you move the wire through the magnetic field (greater speed means greater e.m.f.)
  - if you coil the wire so that more wire is influenced by the magnetic field, you will induce a greater e.m.f.
- As a straight wire moves in a magnetic field it cuts the flux in the area swept out by the wire. The area swept out per second by a wire of length l moving at a velocity $v = lv$. The flux cut per second is therefore Blv. This gives the induced e.m.f. as $\varepsilon = Blv$.
- The laws of electromagnetic induction are:
  - Faraday's law of electromagnetic induction

    The magnitude of an induced e.m.f., ε, is proportional to the rate of change of flux.

    Mathematically, we can write this as $\varepsilon = k\frac{\Delta\Phi}{\Delta t}$

    where k is the constant of proportionality.
  - Lenz's law

    The direction of the induced e.m.f. is such as to oppose the change creating it.

Namibia Not published

- If we include Lenz's law about the direction of the induced e.m.f. with our mathematical expression for Faraday's law we get $\varepsilon = -\frac{\Delta\Phi}{\Delta t}$
  For a coil with N turns this can be written as $\varepsilon = -\frac{N\Delta\Phi}{\Delta t}$ and since $\Phi = BA$ for a coil of area A $\varepsilon = -\frac{N\Delta(BA)}{\Delta t}$
- The direction of the induced e.m.f. can be predicted using the right-hand rule.
- Eddy currents produce a heating effect (since $P = I^2R$). This principle is used in induction welding and in many manufacturing processes.
- Self-inductance is the ratio of the magnetic flux (Φ) times turns of wire (N) to the current.
  Mutual inductance is when a change in the magnetic field due to one coil induces an e.m.f. in another coil.
  Henry is the unit of inductance, 1 H = 1 Wb/A.
- The factors that determine the magnitude of self- and mutual inductance are:
  - magnetic flux
  - number of turns on coil (or coils in mutual inductance)
  - current through the coil (decreasing current increases inductance).
- The inductance of a solenoid is given by $L = n^2\mu_0 A$
- The energy stored in an inductor PE_B is given by $\frac{1}{2}LI^2$
- Magnetic energy density is defined as $\eta_B = \frac{\text{energy}}{\text{volume}}$

Review questions

1. a) Define magnetic flux.
   b) Calculate the magnetic flux when a magnetic field of strength 25 mT passes through an area of 5 cm^2 that is at an angle of 45° to the magnetic field.
2. a) State the laws of electromagnetic induction.
   b) How could you demonstrate an induced e.m.f.?
   c) What rule is used to predict the direction of the induced e.m.f.?
3. a) A wire is coiled into 25 circular coils with a radius of 3 cm. It is then passes it through a magnetic field of strength 10 mT at an angle of 75°. What is the flux linkage?
   b) The coil of wire in part a) is moved from the magnetic field to a place completely outside the magnetic field in a time of 0.5 seconds. What e.m.f. is induced in the coil?

© MoE Not to

c) Calculate the induced e.m.f. in the coil in part a) when the flux density through the coil changes at a rate of 0.6 T/s.

4. a) What are eddy currents and how are they produced?

   b) Give an example of uses of eddy currents.

5. a) What factors determine the magnitude of self and mutual inductance?

   b) Find the self inductance in a coil where the magnetic field is 50 mT, the area of the coil is 10 cm^2, there are 500 turns on the coil and the current through the coil is 2 A.

   c) Find the induced voltage in an inductor of 60 mH when the current is changing at a rate of 1.5 A/s.

6. a) Find the inductance of an air cored solenoid of 250 turns per unit length and area 2.5 cm^2. The permeability of free space is $4\pi \times 10^{-7}$ T m^2/A.

   b) Find the energy stored in the inductor in part (a) when a current of 2.5 A flows through it.

7. a) Define magnetic energy density.

   b) Find the energy density for an inductor with an air core with a magnetic field of 0.5 T. The permeability of free space is $4\pi \times 10^{-7}$ T m^2/A.

7.2 Alternating current (a.c.) generator and transformers

By the end of this section you should be able to:

- Compare direct current (d.c.) and alternating current (a.c.) in qualitative terms.
- Derive the expression for the e.m.f. induced in a rotating coil $\varepsilon = \omega NBA\sin\omega t$.
- Draw a schematic diagram for a simple a.c. generator.
- Explain the working mechanism of a generator.
- Draw a schematic diagram of a transformer.
- Derive transformer equation $\frac{V_1}{V_2} = \frac{N_1}{N_2} = \frac{I_2}{I_1}$ from Faraday's law.
- Explain the importance of alternating current in the transmission of electrical energy.

Direct current (d.c.) and alternating current (a.c.)

Direct current (d.c.) has a constant value. It is the type of current that is obtained from cells that you used in Unit 5. Alternating current (a.c.) is constantly varying. It is the type of current that is transmitted from power stations to consumers. If you were to

© FDRE Ethiopia Not to be republished

look at alternating current on an oscilloscope screen , it would have a sinusoidal waveform like the waves that you met in Unit 2. In constrast, if you were to look at d.c. current on an oscilloscope screen, it would take the form of a straight horizontal line.

Electric generators

An electric generator converts mechanical energy to electrical energy. Figure 7.11 shows a schematic diagram for a simple a.c. generator.

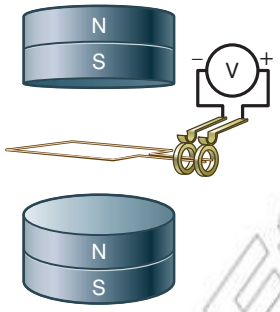


Figure 7.11

We know that a coil in a magnetic field experiences a torque which will make it rotate. As it rotates an e.m.f. is induced. In an a.c. generator the coil is attached to two continuous rings which, in turn, connect in to the external circuit. The magnetic field, B, is constant, and the area, A, of the coil is constant, but the angle between the field and the loop, θ, is changing. We therefore have to use the following expression for flux

$$\Phi = BAN\cos\theta$$

Therefore

$$\frac{\Delta\Phi}{\Delta t} = NAB\frac{\Delta\cos\theta}{\Delta t}$$

If the coil is rotating with a frequency f, then the angle θ is changing as $\theta = 2\pi ft$. The induced voltage is therefore found by using

$$\varepsilon = -\frac{\Delta\Phi}{\Delta t}$$

and substituting $\theta = 2\pi ft$ before differentiating the expression with respect to t.

Thus $\varepsilon = NAB\times 2\pi f\sin(2\pi ft)$

Remember that we can replace $2\pi f$ by ω and so the expression becomes

$$\varepsilon = NAB\times\omega\sin(\omega t)$$

© MoE-EIAR Ethiopia Not to be republished

Activity 7.5: Build a generator

Work in a small group. Wind a long piece of sturdy insulated wire round a soft drink can several times and then remove the can so that you have a wire coil with several loops. Wrap the ends of the wire around the loops so that they do not separate. Leave enough wire at the ends of the coil to be able to bend them into support leads. Remove the insulating material from the leads of the coil. Obtain a metal wire clothes hanger and cut it to form conducting supports for the coil. Stand the supports upright by fitting the ends into holes on a wooden base board.

Predict how the coil will behave when the north pole of a bar magnet is placed near the coil. Use a strong magnet to check your prediction.

Worked example 7.10

A turbine has a coil of area 10 m^2 with 2000 turns rotating in a field of 50 T. The frequency is 0.02 cycles/s. Calculate the induced e.m.f. from the turbine each second.

ε	N	A	B	$2\pi f$	$\sin(2\pi ft)$
?	2000	10	50	0.1256	2.19×10^{-3}

Use $\varepsilon = NAB \times 2\pi f \sin(2\pi ft)$

$= 2000 \times 10 \times 50 \times 0.1256 \times 2.19 \times 10^{-3}$

$= 275.1$ V

DID YOU KNOW?

In Ethiopia, several projects have been initiated to generate more electricity using hydroelectric power. In hydroelectric power, the mechanical energy is supplied by water running through turbines which rotate and generate the electricity.

Transformers

A transformer is used to change voltage. Figure 7.12 shows a diagram of a transformer. You will see that it is very similar to Figure 7.5 on page 273.

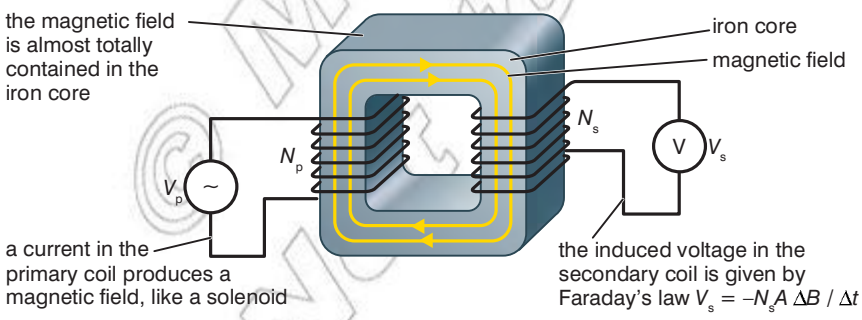


***Figure 7.12** A transformer*

An alternating current (a current that varies with time) is supplied to the primary coil. As this current is constantly changing, a constantly changing magnetic field is set up. This changing field induces an e.m.f. in the secondary coil which, unlike the situation in Figure 7.5, does not decrease to zero, but varies constantly at the same rate as the alternating current supplied to the primary. The magnitude of the induced e.m.f. depends on the strength of the magnetic field (which depends on the number of turns on the primary coil) and the number of turns on the secondary coil. The ratio of the number of turns on the primary coil to the number of turns on the secondary coil is the same as the ratio of the voltage on the primary coil to the voltage on the secondary coil

$$\frac{N_p}{N_s} = \frac{V_p}{V_s}$$

This relationship enables transformers to change voltage, for example, to decrease the mains supply down to a level suitable for use in an appliance.

Worked example 7.11

A transformer in the electricity supply network changes a voltage from 11 kV to 415 V. The primary coil in this transformer has 8000 turns. How many turns will be needed on the secondary coil to give the correct output voltage?

N_p	N_s	V_p (V)	V_s (V)
8000	?	11×10^3	415

Use

$$\frac{N_p}{N_s} = \frac{V_p}{V_s}$$

$$N_s = \frac{N_p V_s}{V_p}$$

$$= \frac{8000 \times 415}{11 \times 10^3}$$

$$= 302$$

Activity 7.6: Investigating the uses of transformers

Carry out some research into the uses of transformers. Present your findings to your class in the form of a report.

The transmission of electrical energy

After electricity has been generated at a power station, it needs to be transmitted through cables to consumers such as businesses and homes. The electricity generated by the power station is in the form of alternating current.

When electricity flows through cables, some power is lost since power = I^2R where I is the current and R is the resistance of the cables. Clearly, in an electricity distribution system the power loss

© MoE Ethiopian Not to be republished

needs to be minimized. This is done by transmitting the electricity at very high voltages (above 110 kV or above) since a higher voltage reduces the current. For example, increasing the voltage by a factor of 10 reduces the current by a factor of 10 and therefore the energy lost by a factor of $10^2 = 100$. However, consumers need the electricity to reach their business or home at a much lower voltage – in Ethiopia the mains voltage is 220 V.

When alternating current is used to transmit electricity, transformers can be used to increase or decrease the voltage as required, as explained in the previous section. However, if direct current were to be used to transmit electricity, it is not as straightforward to increase or decrease the voltage as required. For this reason, electricity is only transmitted in the form of direct current over extremely long distances (over about 30 km where alternating current can no longer be applied). At these distances, the cost of converters (from a.c. to d.c. and back again) at each end of the line is offset by the lower cost of construction and lower energy losses.

Activity 7.7: Investigating the national grid set by the Ethiopian Electric Power Corporation (EEPCo) and a.c.

Work with a partner. Carry out some research into the national grid in Ethiopia and how it uses alternating current. Present your findings to your class in a form of your choice.



© FDRE Ministry of Education Not to be republished

Summary

In this section you have learnt that:

- Direct current (d.c.) is current that has a constant value.
- Alternating current (a.c.) varies continuously with time – on an oscilloscope screen it is a sinusoidal waveform.
- The e.m.f. induced in a rotating coil $\varepsilon = \omega NAB\sin\omega t$
- A schematic diagram for a simple a.c. generator is as follows:

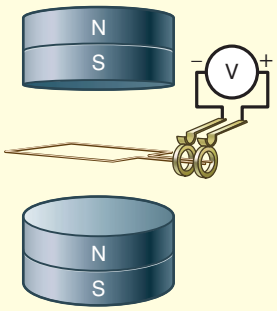


Figure 7.13

- A coil in a magnetic field experiences a torque which will make it rotate. As it rotates an e.m.f. is induced. In an a.c. generator the coil is attached to two continuous rings which, in turn, connect in to the external circuit.
- A schematic diagram of a transformer is as follows:

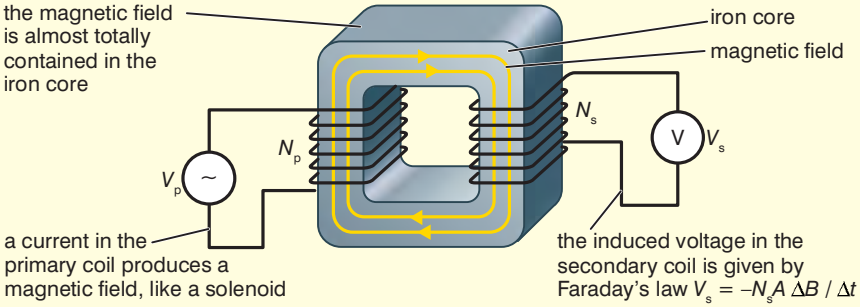


Figure 7.14

- The transformer equation is $\frac{V_1}{V_2} = \frac{N_1}{N_2} = \frac{I_2}{I_1}$ and can be derived from Faraday's law.
- The electricity generated by the power station is in the form of alternating current.

Review questions

1. Draw an oscilloscope trace for
   a) alternating current
   b) direct current.
2. Draw a simple schematic diagram of an a.c. generator and explain how it works.
3. a) Draw a simple schematic diagram of a transformer and explain how it works.
   b) A transformer converts the mains supply from 220 V down to 2 V. There are 160 turns on the secondary coil. How many turns does the primary coil have?
4. What is the main reason why alternating current is used to transmit electricity?

© MoE Not to be

7.3 Alternating current (a.c.)

- Explain what is meant by r.m.s. values.
- Apply the relationship between r.m.s. and peak values for the current and potential difference for a sinusoidal waveform.
- Identify that the current and voltage are in phase in a resistor in an a.c. circuit.
- Explain the behaviour of a capacitor in an a.c. circuit.
- Derive the expression for the instantaneous current and voltage in a resistive and capacitive circuit.
- Identify that the current leads the voltage by $\frac{\pi}{2}$ in a capacitor in an a.c. circuit.
- Draw phasor diagrams for resistive and capacitive circuits.
- Define capacitive reactance.
- Use the terms: r.m.s. current, r.m.s. potential difference, peak current, peak potential difference, half cycle average current, phase difference, phase lag, phase lead.
- Use the terms: reactance, impedance, power factor with their correct scientific meaning.
- Define the power factor in an a.c. circuit.
- Identify that the voltage leads the current by $\frac{\pi}{2}$ in an inductive circuit.
- Explain the behaviour of an inductor in an a.c. circuit.
- Derive the expression for the instantaneous current/voltage in an inductor in an inductive circuit.
- Define inductive reactance.
- Describe the behaviour of an RL circuit.
- Describe the behaviour of an LC circuit.
- Describe the behaviour of RLC circuits.
- Derive an expression for the impedance of RLC circuits.
- Draw phasor diagrams for an RLC circuit.
- Solve problems involving the magnitude and phase of current and applied p.d. in a.c. circuits which include resistors, capacitors and inductors.

Ethiopia published

KEY WORDS

root mean square (r.m.s.) value *a value for the current or voltage that would be equivalent to the effective steady value*

half cycle average current *is given by the relation* $I_{avg} = 0.637 \times I_p$ *where* I_p *is the peak current.*

peak current *the maximum value of the current in a cycle*

peak potential difference *the maximum value of the voltage in a cycle*

r.m.s. current *the value of the current that would be equivalent to the effective steady value*

r.m.s. potential difference *the value of the potential difference that would be equivalent to the effective steady value*

Root mean square (r.m.s.) values in a.c. circuits

Alternating current has a sinusoidal waveform, as shown in Figure 7.15.

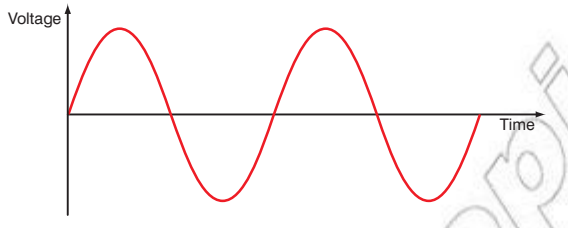


Figure 7.15

This means that its magnitude is varying continuously but its average magnitude is zero. The **half cycle average current** is given by the relation

$I_{avg} = 0.637 \times I_p$

where I_p is the peak value of the current.

In order to analyse a.c. circuits, we need to use a value for the current or voltage that would be equivalent to the effective steady value. This value is called the **root mean square (r.m.s.) value**.

Squaring a number always gives a positive result. So, we should be able to find a real average of a squared value. Then we take the square root of that average to find the effective current or voltage.

We use the following relationships for sinusoidal waveforms.

$I_{rms} = \frac{I_{peak}}{\sqrt{2}} = I_{\text{peak}} \times 0.707$ where I_{rms} is the **r.m.s. current** and I_{peak} is the maximum value of the current in a cycle (the **peak current**)

$V_{rms} = \frac{V_{peak}}{\sqrt{2}} = V_{\text{peak}} \times 0.707$ where where V_{rms} is the **r.m.s. potential difference** and V_{peak} is the maximum value of the potential difference in a cycle (the **peak potential difference**)

© MoE, EPDI Ethiopia Not to be republished

Worked example 7.12

An alternating supply has a peak value of 2.5 A for the current and a peak value of 6 V for the voltage. Find the r.m.s. value for a) the current b) the voltage from this supply.

a)

I_{peak} (A)	I_{rms} (A)
2.5	?

Use $I_{rms} = \frac{I_{peak}}{\sqrt{2}} = I_{peak} \times 0.707$

$I_{rms} = 2.5 \times 0.707$

$= 1.7675$ A

b)

V_{peak} (V)	V_{rms} (V)
6	?

Use $V_{rms} = V_{peak} \times 0.707$

$V_{rms} = 6 \times 0.707$

$= 4.242$ V

Resistive circuits and alternating currents

Consider the circuit shown in Figure 7.16.

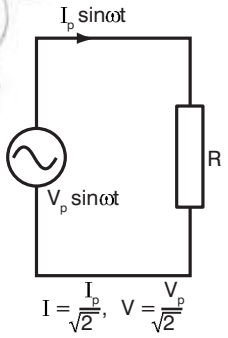


Figure 7.16

If we use the r.m.s. values for the current and voltage through the resistor, then for ordinary currents and frequencies, the behaviour of a resistor in an a.c. circuit is the same as the behaviour of a resistor in a d.c. circuit.

We know that voltage and current from an a.c. supply are both sinusoidal waveforms such as those you met in Unit 2. You know that it is possible for two such waveforms to be in phase with each other (peaks and troughs at same instant) or out of phase (peaks and troughs at different times) or in antiphase (peak on one wave at the same time as a trough on the other). When we consider a.c. circuits we need to know whether the current and voltage are in phase with each other or out of phase. In a resistor, the voltage and current increase and decrease directly with each other, as you would expect from Ohm's law. We say that they are in phase, as shown in Figure 7.17.

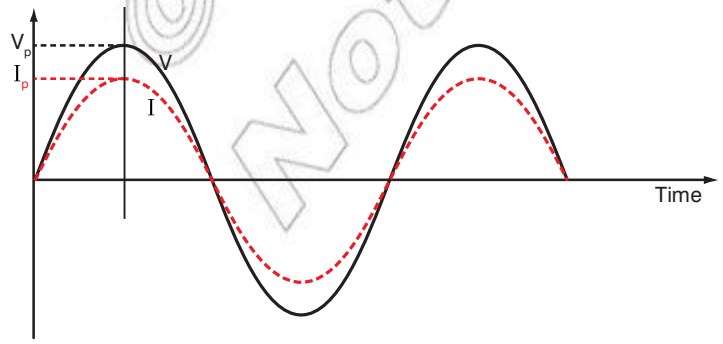


Figure 7.17

© Ministry of Education Not to be republished

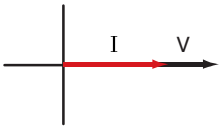


Figure 7.18

This can also be represented using a phasor diagram like the one shown in Figure 7.18.

Activity 7.8: Demonstrating that current and voltage are in phase in a resistive a.c. circuit

Set up a circuit as shown in Figure 7.16. Use a signal generator as the a.c. supply. Connect an oscilloscope so that it shows you the current and voltage in the circuit. You should see a trace on the screen which resembles that in Figure 7.17.

Capacitive circuits and alternating currents

Consider the circuit shown in Figure 7.19.

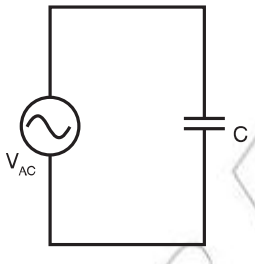


Figure 7.19

The capacitor will draw current to oppose any change of voltage across itself. To find out how much current it will draw, we need to go back to some capacitor basics. In Unit 4, you learnt that

$$Q = CV$$

where Q is the charge on the capacitor, C is the capacitance and V is the p.d. across the capacitor.

You also know that

$$I = \frac{\Delta\Phi}{\Delta t}$$

In this circuit the charging current changes constantly as the voltage across C changes. The value of C is constant so we can combine the two equations to give

$$I = C\frac{\Delta\Phi}{\Delta t}$$

When we use an alternating voltage source, the voltage is a sine wave of some frequency, f. Mathematically, we write

$$V_c = V_p\sin(2\pi ft) = V_p\sin(\omega t)$$

where V_c is the p.d. across the capacitor, V_p is the peak value of the p.d., and $\omega = 2\pi f$.

To find the current, we need to find the derivative of $V_p\sin(\omega t)$ with respect to t. To do this, we use the standard mathematical result shown in the box.

Therefore the current across a capacitor in an a.c. circuit is given by

$$I = CV_p\omega\cos(\omega t) = \omega CV_p\cos(\omega t)$$

The derivative of sin(ax)

Mathematical formulae books give the following result:

$$\frac{d(\sin(ax))}{dx} = a\cos(ax)$$

Not to be republished © ETDP Ethiopia

This equation tells us that the current resulting from applied a.c. voltage (which is a sine wave) is shifted in phase by $\frac{\pi}{2}$ as shown in Figure 7.20. The applied a.c. voltage is shown in red and the resulting current is shown in blue. There is a **phase difference** between the current and the applied voltage. There is a **phase lead** between the current and the applied voltage of $\frac{\pi}{2}$.

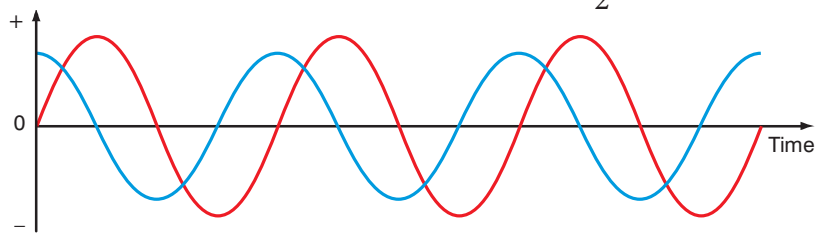


Figure 7.20

This fits in with what we know about the capacitor, which is that it will draw current in its attempt to oppose any change of voltage across its terminals.

We can draw a phasor diagram to represent this as shown in Figure 7.21.

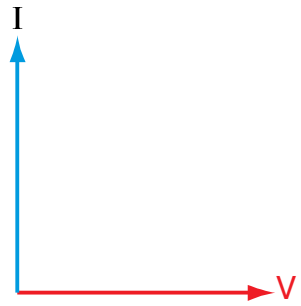


Figure 7.21

The factor ωC (or $2\pi fC$) is a constant of proportionality that depends on both the value of C and the frequency of the sine wave and relates the voltage and current in a capacitor. As either the frequency or the value of C increases, the capacitor current will increase for the same applied voltage. If we compare this to resistance, that relates voltage and current in a resistor, we see that this is the opposite behaviour to a resistor, where as the value of R increases for the same applied voltage, the value of the current will decrease.

If we invert this factor and use the factor $\frac{1}{\omega C}$ then it will behave like the capacitive equivalence of resistance. We cannot call it resistance but, because the capacitor reacts to the applied voltage, we call it the **reactance**. Reactance is measured, like resistance, in ohms and is generally given the symbol X and capacitive reactance is given the symbol X_c.

$$X_c = \frac{1}{2\pi fC} = \frac{1}{\omega C}$$

KEY WORDS

phase difference *difference in phase between two sine waves*

phase lead *where one sine wave leads another by a given number of degrees*

reactance *the equivalent quantity to resistance when we are talking about capacitors or inductors*

© Wced FDR Ethics not to be republished

Activity 7.9: Demonstrating the phase difference between the current and voltage in a capacitive a.c. circuit

Set up a circuit as shown in Figure 7.19. Use a signal generator as the a.c. supply. Connect an oscilloscope so that it shows you the current and voltage in the circuit. You should see a trace on the screen which resembles that in Figure 7.20.

Worked example 7.13

An alternating supply has a frequency of 50 Hz (50 cycles/s). A capacitor is connected in a circuit with this alternating supply as shown in Figure 7.19. The value of the capacitor is 1 μF. What is the reactance of this capacitor?

X_c (Ω)	C (F)	f (Hz)
?	1×10^{-6}	50

Use

$$X_c = \frac{1}{2\pi fC}$$

$$= \frac{1}{2 \times \pi \times 50 \times 1 \times 10^{-6}}$$

$$= 3184.7\ \Omega$$

Inductive circuits and alternating currents

Consider the circuit shown in Figure 7.22.

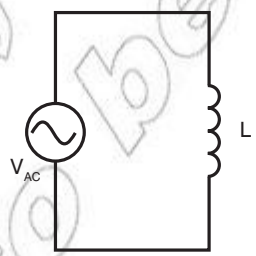


Figure 7.22

To explain the behaviour of the inductor in this circuit, we need to find the expression for the current through the inductor and the p.d. across the conductor, as we did for the capacitor in a similar circuit.

We know that, for an inductor, the following relationship is true

$$\varepsilon = L\frac{\Delta I}{\Delta t}$$

As we are considering alternating sources of p.d., ε is a sine wave, described mathematically as

$\varepsilon = V_p \sin(\omega t)$ where $\omega = 2\pi f$.

© MoE, EV not to be

This gives us

$$V_p \sin(\omega t) = L\frac{\Delta I}{\Delta t}$$

If we compare this to the expression we get from Ohm's law

$$\frac{V}{R} = I$$

we can see that we can define an inductive reactance as

$$X_L = \omega L = 2\pi f L$$

From the expression for the current through the inductor, we can see that the current is a negative cosine wave and so there will be a **phase lag** of $\frac{\pi}{2}$ between the current and the applied p.d., as shown in Figure 7.23 (again, the p.d. is in red and the current in blue).

KEY WORDS

phase lag *when one sine wave lags behind another sine wave*

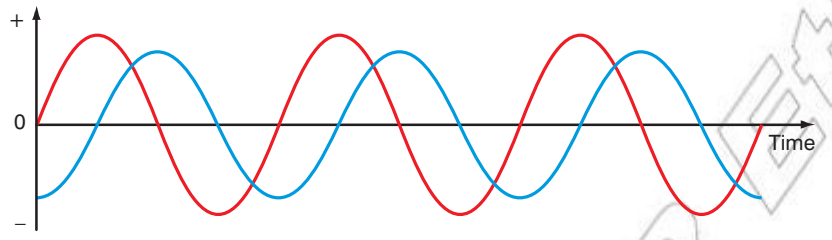


Figure 7.23

This seems reasonable since we know that the reaction of any inductor is to oppose any change of current through itself. We can draw a phasor diagram to show this relationship as shown in Figure 7.24.

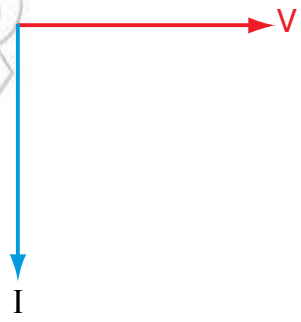


Figure 7.24

Activity 7.10: Demonstrating the phase difference between the current and voltage in an inductive a.c. circuit

Set up a circuit as shown in Figure 7.22. Use a signal generator as the a.c. supply. Connect an oscilloscope so that it shows you the current and voltage in the circuit. You should see a trace on the screen which resembles that in Figure 7.23.

Worked example 7.14

An alternating supply has a frequency of 50 Hz (50 cycles/s). An inductor is connected in a circuit with this alternating supply as shown in Figure 7.22. The inductance of the inductor is 1 μH. What is the reactance of this inductor?

X_L (Ω)	L (H)	f (Hz)
?	1×10^{-6}	50

Use $X_L = 2\pi f L$

$= 2 \times \pi \times 50 \times 1 \times 10^{-6}$

$= 3.14 \times 10^{-4}\ \Omega$

FOR Ethiopia republished

By comparing the answers to worked examples 7.13 and 7.14, you can see that a capacitor offers high reactance in a circuit and an inductor offers low reactance in a circuit. We shall now explore what happens when we combine components in a circuit.

KEY WORDS

power factor *The power factor in an a.c. circuit is defined as power factor = cos θ where θ is the phase difference between the p.d. and the current*

impedance *the total opposition to the flow of current in an a.c. circuit*

The power factor in an a.c. circuit

The **power factor** in an a.c. circuit is defined as

$$\frac{\text{real power flowing through the load}}{\text{apparent power}}$$

Real power is the capacity of a circuit for performing work at a particular time. Apparent power is the product of the current and p.d. of the circuit.

When there is a load such as a capacitor or an inductor in an a.c. circuit, energy stored in the loads results in a time difference between the p.d. and the current (the phase difference). During each cycle of a.c., extra energy, in addition to the energy consumed in the load, is temporarily stored in electric or magnetic fields, and then returned to the circuit later in the cycle.

If the phase angle between the current and the p.d. is θ then the power factor is given by cos θ.

Combining a resistor and an inductor in a circuit

You know that the p.d. in an inductive circuit leads the current. When there is another component in the circuit, such as a resistor as shown in Figure 7.25, the phase difference between the p.d. and the current is not $\frac{\pi}{2}$, but an angle θ.

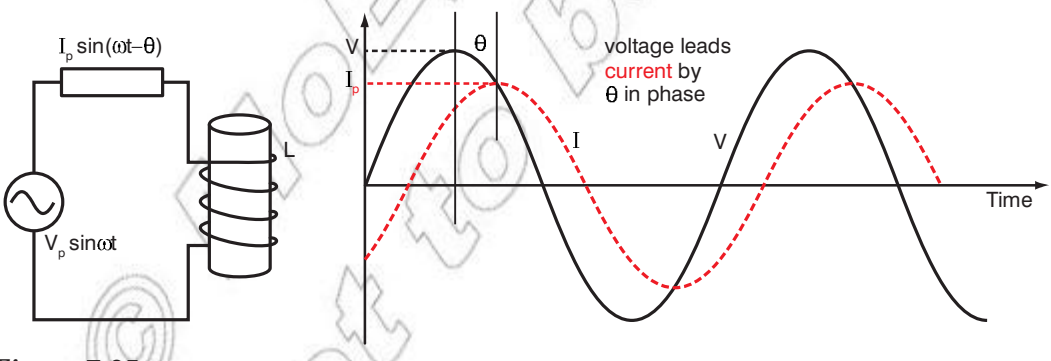


Figure 7.25

The total opposition to the flow of current is a combination of the resistance from the resistor and reactance from the inductor. We cannot use either term in this case. We use the term **impedance** to describe the total opposition to the flow of current in circuits which combine resistors and inductors (or, as we shall see later, capacitors). Impedance is generally given the symbol *Z*.

To analyse an a.c. circuit containing different components, we generally use complex numbers. The basic mathematical ideas you need to understand the rest of this section are given in the box.

© NOE Not to be copied

We can write the total impedance in the circuit shown in Figure 7.23 as

$Z = R + j\omega L$

since the components are in series.

This means that we can find the magnitude of Z using

$Z = \sqrt{R^2 + \omega^2 L^2}$

and the phase difference θ using

$\theta = \tan^{-1}\frac{\omega L}{R}$

We can draw a **phasor diagram** for the circuit as shown in Figure 7.27.

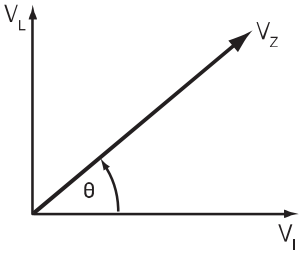


Figure 7.27

You can apply Ohm's law to such circuits as long as you use the value of the impedance of the circuit. The formula then becomes

$V = IZ$

Worked example 7.15

An alternating supply of 12 V and frequency 50 Hz is used in a circuit which has a resistor of 100 Ω and an inductor of 30 mH in series. Calculate:

a) the total impedance of the circuit

b) the current in the circuit

c) the phase angle between the supply and the current

d) the power factor for the circuit.

a)

Z (Ω)	R (Ω)	ω (Hz)	L (H)
?	100	$2\pi \times 50$	30×10^{-3}

Use

$Z = \sqrt{R^2 + \omega^2 L^2}$

$= \sqrt{(100)^2 + (2\pi \times 50)^2 (30 \times 10^{-3})^2}$

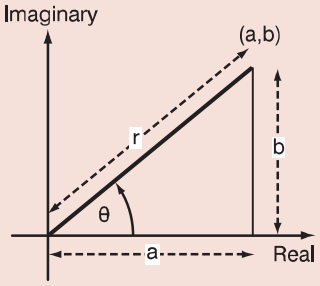
$= \sqrt{10\,000 + 98\,596 \times (9 \times 10^{-4})}$

$= \sqrt{10\,088.7}$

$= 100.44\ \Omega$

Complex numbers

A complex number takes the form $a + jb$ where a is the real part of the number and b is the imaginary part of the number. We can plot the complex number $a + jb$ on an Argand diagram, as shown in Figure 7.26.



the point (a,b) represents the complex number a + jb

Figure 7.26

From Figure 7.26, you can see that the magnitude, r, of the complex number can be found by applying Pythagoras' theorem

$r = \sqrt{a^2 + b^2}$

and that the angle θ can be found using

$\tan\theta = \frac{b}{a}$

so that

$\theta = \tan^{-1}\frac{b}{a}$

b) Use Ohm's law

$$I = \frac{V}{Z}$$

$$= \frac{12}{100.44}$$

$$= 0.119 \text{ A}$$

c)

θ	ω (Hz)	L (H)	R (Ω)
?	$2 \times \pi \times 50$	30×10^{-3}	100

Use

$$\theta = \tan^{-1} \frac{\omega L}{R}$$

$$= \tan^{-1} \frac{2 \times \pi \times 50 \times 30 \times 10^{-3}}{100}$$

$$= \tan^{-1} \frac{9.42}{100}$$

$$= \tan^{-1} 0.0942 = 5.38°$$

d) power factor = $\cos \theta = 0.996$

Activity 7.11: Investigating an LR circuit

Set up the circuit shown in Figure 7.28. Investigate how the p.d. across the inductor varies with time by using an oscilloscope to display the p.d. How does varying the value of R alter the p.d. across the inductor?

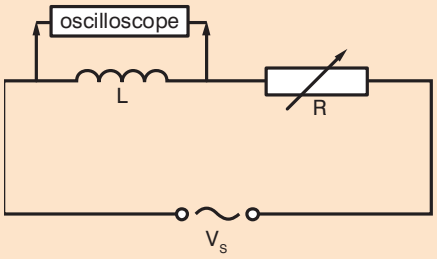


Figure 7.28

Write a report which describes what you observe.

Activity 7.12: Investigating inductors in stage lighting

Work in a small group to investigate and report on the uses of inductors in dimmer switches in stage lighting.

© MoE Not to

Combining a resistor and a capacitor in a circuit

Consider the circuit shown in Figure 7.29.

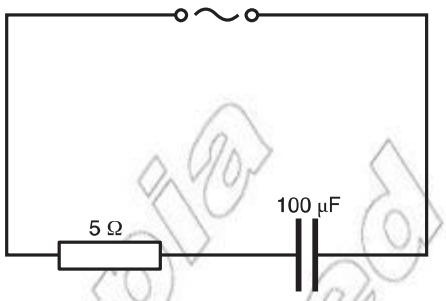


Figure 7.29

We can analyse this circuit and find the impedance, the current through the circuit, the phase angle and the power factor as we did for the circuit which combined an inductor and a resistor.

The impedance, Z, is given by

$$Z = \sqrt{R^2 + (X_c)^2}$$

and the phase angle θ is given by

$$\theta = \tan^{-1}\frac{X_c}{R}$$

Worked example 7.16

An alternating supply of 12 V and frequency 50 Hz is used in a circuit which has a resistor of 5 Ω and a capacitor of 100 μF in series. Calculate

a) the total impedance of the circuit

b) the current in the circuit

c) the phase angle between the supply and the current

d) the power factor for the circuit

a) First we need to find Xc

X_c (Ω)	f (Hz)	C (F)
?	50	1×10^{-6}

Use

$$X_c = \frac{1}{2\pi fC}$$

$$X_c = \frac{1}{2\pi \times 50 \times 1 \times 10^{-6}}$$

$$= \frac{1}{3.14 \times 10^{-6}}$$

$$= 3185\ \Omega$$

Z (Ω)	R (Ω)	X_c (Ω)
?	5	3185

Use

$$Z = \sqrt{5^2 + (3185)^2}$$

$$= \sqrt{25 + 10\,144\,225}$$

$$= \sqrt{10\,144\,250}$$

$$= 3185\ \Omega$$

Note that the contribution of the resistor to the impedance compared to the contribution of the capacitor is negligible.

b) Use Ohm's law

$$I = \frac{V}{Z}$$

$$= \frac{12}{3185}$$

$$= 3.77\ \text{mA}$$

c) Use

$$\theta = \tan^{-1}\frac{X_c}{R}$$

$$= \tan^{-1}\frac{3185}{5}$$

$$= 89.91°$$

d) The power factor is given by $\cos\theta = 1.57 \times 10^{-3}$

Activity 7.13: Investigating an RC circuit

Set up the circuit shown in Figure 7.30.

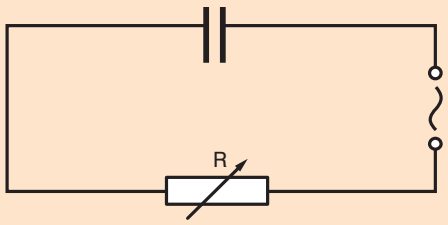


Figure 7.30

Investigate how the p.d. across the capacitor varies with time by using an oscilloscope to display the p.d. Plot a graph to show this relationship.

Change the value of the variable resistor and observe how that affects the result.

You learnt in earlier units that the time constant in an RC circuit is given by multiplying the value of the capacitor by the value of the resistor. Does this relationship hold for your circuit? Write a report which describes what you observe.

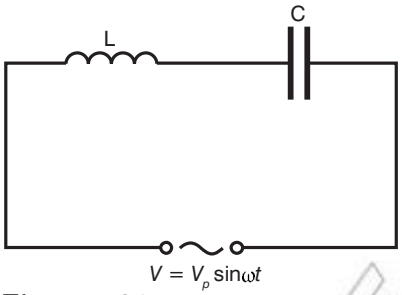


Figure 7.31

Combining an inductor and a capacitor in a circuit

Consider the circuit shown in Figure 7.31.

When the power supply is connected, there will be oscillations between the inductor and the capacitor. We know that the phasor diagrams for an inductor in a circuit and a capacitor in a circuit are as shown in Figure 7.32.

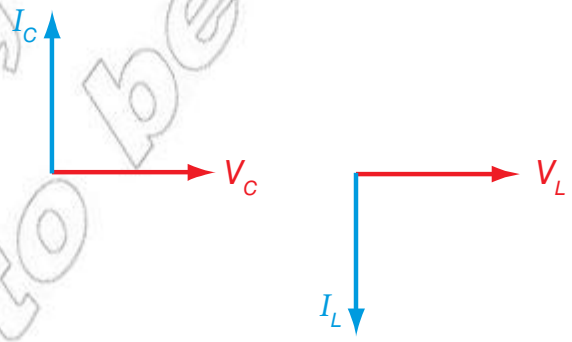


Figure 7.32

In the circuit shown in Figure 7.31, the current through the capacitor, I_C, is the same as the current through the inductor, I_L. We can combine the phasor diagrams as shown in Figure 7.33.

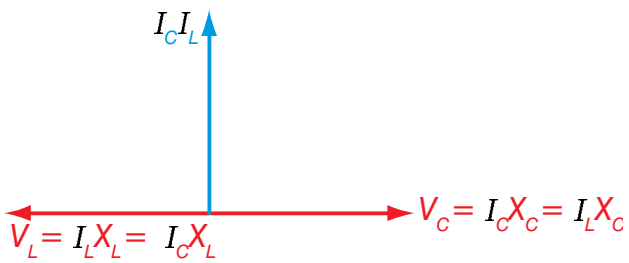


Figure 7.33

© MoE. Not to be reproduced

Since V_L and V_C are in opposite directions, the impedance of this circuit is given by

$$Z = \sqrt{(X_L)^2 - (X_c)^2}$$

There will therefore by a frequency, f, at which $X_L = X_C$ and the impedance will be zero. This will happen when

$$2\pi fL = \frac{1}{2\pi fC}$$

The value of f will be given by

$$f = \frac{1}{2\pi\sqrt{LC}}$$

This value of f is called the resonant frequency of the circuit and the circuit will conduct extremely well at this frequency. In practice, resistors are often added to such circuits in order to damp the oscillations – we shall learn more about this in the next section.

DID YOU KNOW?

Radios use resonance to tune the receiving circuitry to the broadcast frequency of the station being tuned in. Each radio station broadcasts at a precise carrier frequency. When a receiver is in resonance at this frequency it is in tune with that station. Tuning, in most radios, is done by changing the capacitance of the receiving circuit at fixed inductance.

Activity 7.14: Investigating an LC circuit

Work in a small group. Set up a circuit as shown in Figure 7.31.

Investigate the behaviour of the circuit with different values for L and C and use a signal generator as the a.c. supply so that you can see the effect of changes in frequency. Can you find the resonant frequency for a circuit? Does the frequency you find agree with the theoretical value given above?

Write a report on your observations.

Worked example 7.17

Find the resonant frequency for a circuit containing a 30 μH conductor and a 2.0 pF capactor.

f (Hz)	L (H)	C (F)
?	30×10^{-6}	2×10^{-12}

Use

$$f = \frac{1}{2\pi\sqrt{LC}}$$

$$= \frac{1}{2\pi\sqrt{30 \times 10^{-6} \times 2 \times 10^{-12}}}$$

$$= \frac{1}{4.86 \times 10^{-8}}$$

$$= 2.058 \times 10^{7} \text{ Hz}$$

Combining resistors, capacitors and inductors

Consider the circuit shown in Figure 7.34.

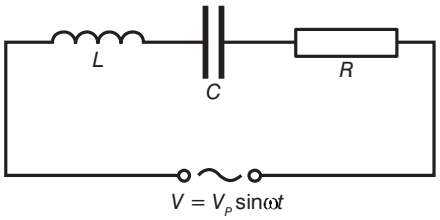


Figure 7.34

We can draw a phasor diagram for this circuit as shown in Figure 7.35.

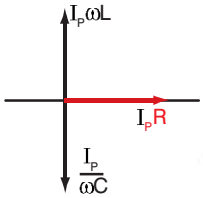


Figure 7.35

The peak potential differences for a circuit such as the one shown in Figure 7.34 are given by

$V_R = I_{peak}R$

$V_L = I_{peak}X_L$

$V_C = I_{peak}X_C$

The sum of the potential differences across the circuit may be written as

$$V_T = \sqrt{(V_R)^2 + (V_L - V_C)^2}$$
$$= \sqrt{(I_{peak}R)^2 + (I_{peak}X_L - I_{peak}X_C)^2}$$
$$= I_{peak}\sqrt{R^2 + (X_L - X_C)^2} = I_{peak}Z$$

Therefore the impedance of the circuit may be written as

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

The phase angle between the current and the p.d. for the circuit is given by

$$\tan\theta = \frac{X_L - X_C}{R}$$

When $X_L > X_C$ (this is generally at high frequencies) the phase angle is positive so the current lags behind the applied p.d. When $X_L < X_C$ the phase angle is negative and the current leads the applied p.d. When $X_L = X_C$ the phase angle is zero and the reactance in the circuit matches the resistance. This is when the resonant frequency is reached. The resonant frequency is given by

$$f = \frac{1}{2\pi\sqrt{LC}}$$

as we found for circuits with an inductor and a capacitor. However, the presence of the resistor will dampen the resonant oscillations.

© MoE-FDRE Ethiopia Not to be republished

Activity 7.15:Investigating an RLC circuit

Work in a small group. Set up a circuit as shown in Figure 7.34.

Investigate the behaviour of the circuit with different values for *R, L* and *C* and use a signal generator as the a.c. supply so that you can see the effect of changes in frequency. Can you find the resonant frequency for a circuit? Does the frequency you find agree with the theoretical value given above? How does changing the value of *R* affect the dampening of the oscillations?

Write a report on your observations.

Worked example 7.18

Consider the circuit shown in Figure 7.36. The alternating supply frequency is 50 Hz.

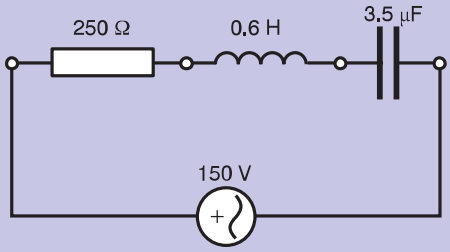


Figure 7.36

Calculate a) the impedance for this circuit b) the phase angle for this circuit c) whether the current or the applied p.d. leads and therefore whether the circuit is predominantly capacitive or predominantly inductive d) the resonant frequency for this circuit.

a)

$Z\ (\Omega)$	$R\ ((\Omega)$	$X_L = 2\pi fL\ (\Omega)$	$X_C = \frac{1}{2\pi fc}\ (\Omega)$
?	250	$2 \times \pi \times f \times L = 188.4$	$\frac{1}{2 \times \pi \times f \times c} = 910$

Use

$$Z = \sqrt{250^2 + (188.4 - 910)^2}$$
$$= \sqrt{62\,500 + 520\,707}$$
$$= \sqrt{583\,207}$$
$$= 764\ \Omega$$

b)

$\tan\theta$	$X_L\ (\Omega)$	$X_C\ (\Omega)$	$R\ (\Omega)$
?	188.4	910	250

Use

$$\tan\theta = \frac{X_L - X_C}{R}$$
$$= \frac{188.4 - 910}{250}$$
$$= -2.8864$$
$$\theta = -70.9°$$

c) The current leads the applied p.d. so the circuit is predominantly capacitive.

d)

f (Hz)	L (H)	C (F)
?	0.6	3.5×10^{-6}

Use

$$f = \frac{1}{2\pi\sqrt{LC}}$$

$$= \frac{1}{2\times\pi\sqrt{0.6 \times 3.5 \times 10^{-6}}}$$

$$= \frac{1}{2\times\sqrt{2.1 \times 10^{-6}}}$$

$$= \frac{1}{6.28 \times 1.45 \times 10^{-3}}$$

$$= \frac{1}{9.106 \times 10^{-3}}$$

$$= 110 \text{ Hz}$$

Summary

In this section you have learnt that:

- The **half cycle average current** is given by the relation
  $I_{avg} = 0.637 \times I_p$
  where I_p is the peak value of the current.
- The **half cycle average p.d.** is given by the relation
  $V_{avg} = 0.637 \times V_p$
  where V_p is the peak value of the p.d.
- In order to analyse a.c. circuits, we need to use a value for the current or voltage that would be equivalent to the effective steady value. This value is called the **root mean square (r.m.s.) value**.
- We use the following relationships for sinusoidal waveforms:

  $$I_{rms} = \frac{I_{peak}}{\sqrt{2}} = I_{peak} \times 0.707$$

  where I_{rms} is the **r.m.s. current**

  and I_{peak} is the maximum value of the current in a cycle (the **peak current**)

  $$V_{rms} = \frac{V_{peak}}{\sqrt{2}} = V_{peak} \times 0.707$$

  where where V_{rms} is the **r.m.s. potential difference** and V_{peak} is the maximum value of the potential difference in a cycle (the **peak potential difference**)
- The current and voltage are in phase in a resistor in an a.c. circuit.
- A capacitor in an a.c. circuit impedes the changing current.
- The expression for the instantaneous current and voltage in a resistive circuit is given by Ohm's law, $V = IR$. The expression for the voltage is $V = V_p\sin(\omega t)$.
- The expression for the instantaneous current in a capacitive circuit is given by $I = CV_p\omega\cos(\omega t) = \omega CVp\cos(\omega t)$.
- The expression for the voltage is given by $V = V_p\sin(\omega t)$.
- The current leads the voltage by $\frac{\pi}{2}$ in a capacitor in an a.c. circuit.
- Phasor diagrams for resistive and capacitive circuits are as follows.

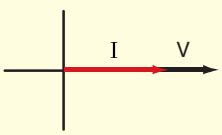


Figure 7.37

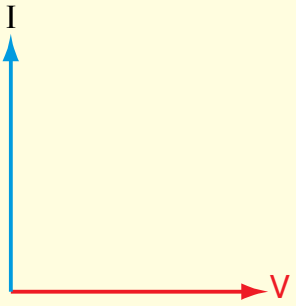


Figure 7.38

- Capacitive reactance is the amount by which the capacitor impedes the flow of current and is given by the expression.
  $X_C = \frac{1}{2\pi fC}$ where f is the frequency of the supply p.d. and C is the value of the capacitance.
- **Phase difference** is the difference in phase between two sine waves.
- **Phase lead** is where one sine wave leads another by a given number of degrees.
- Phase lag is where one sine wave lags behind another by a given number of degrees.
- **Reactance** is the equivalent quantity to resistance when we are talking about capacitors or inductors.
- **Impedance** is the total opposition to the flow of current in an a.c. circuit.
- **The power factor** in an a.c. circuit is defined as
  $$\frac{\text{real power flowing through the load}}{\text{apparent power}} = \cos\theta$$
  where θ is the phase difference between the p.d. and the current.
- The voltage leads the current by $\frac{\pi}{2}$ in an inductive circuit.
- An inductor in an a.c. circuit impedes the flow of current.
- The expression for the instantaneous current in an inductor in an inductive circuit is
  $-\frac{V_p\cos(\omega t)}{\omega L} = I$
- Inductive reactance is the amount by which an inductor impedes the flow of current and is given by the expression $X_L = 2\pi fL$ where f is the frequency of the supply p.d. and L is the value of the inductance.
- LC and RLC circuits have a resonant frequency, given by
  $f = \frac{1}{2\pi\sqrt{LC}}$
  at which the impedance is very low and so it is easy for current to flow through the circuit.
- An expression for the impedance of RLC circuits is
  $Z = \sqrt{R^2 + (X_L - X_C)^2}$
- A phasor diagram for an RLC circuit is as shown in Figure 7.39.

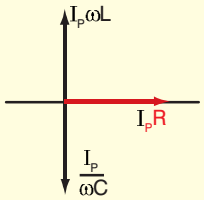


Figure 7.39

© Not

Review questions

1. Explain why, if you pass a steady 1 A d.c. current through a bulb, then pass an alternating current whose root mean square value is 1 A through the bulb, the bulb glows equally brightly on both occasions.
2. An alternating supply has an r.m.s. value of 12 V. Calculate
   a) its peak value
   b) its half cycle average value.
3. Calculate the reactance of a 100 μF capacitor at
   a) 50 Hz
   b) 1000 Hz
   c) Why do these values differ?
4. An inductor of 0.8 H is connected in series with a 100 Ω resistor. It is connected to an alternating supply of 12 V at 50 Hz.
   a) Calculate the reactance of the inductor.
   b) Calculate the impedance of the circuit.
   c) Work out the current drawn from the supply.
   d) i) Find the phase angle between the current and the voltage.
      ii) Which is ahead – the current or the voltage?
      iii) Draw a phasor diagram to illustrate this.
   e) Find the power factor for this circuit.
5. A capacitor of 100 μF is added to the circuit in question 4. Find the resonant frequency for the circuit.

7.4 Power in a.c. circuits

By the end of this section you should be able to:

- Show that the average power in an a.c. capacitive circuit is zero.
- Derive the expression for the average power in an a.c. inductive circuit.
- Derive the expression for the average power in an a.c. RLC circuit.
- Distinguish between real, apparent and ideal power of an RLC circuit.

© MoE FDR Ethiopia Not to be republished

Average power in a.c. capacitive and inductive circuits

In an electrical circuit, energy is supplied by the supply p.d., stored by capacitive and inductive elements and dissipated by resistive elements.

The principle of conservation of energy means that, at any time t, the rate at which energy is supplied by the supply p.d. must equal the sum of the rate at which it is stored in the capacitive and inductive elements and dissipated by the resistive elements (here we assume that ideal capacitors and inductors have no internal resistance).

We know that supply p.d. is given by

$V = V_p \sin(\omega t)$

where V_p is the peak value of the p.d.

We also know that power $= \frac{\text{work done}}{\text{time}} = \frac{\Delta W}{\Delta t}$

In the case of electrical energy, we also know that

power = p.d. × current

In an a.c. circuit, we know that the current is given by

$I = I_p \sin(\omega t - \theta)$

where I_p is the peak value for the current and θ is the phase angle for the circuit.

We can equate the two expressions for power to give the power at any time, t

$$\frac{\Delta W}{\Delta t} = V_p \sin(\omega t) \times I_p \sin(\omega t - \theta)$$
$$= V_p I_p \sin(\omega t)[\sin(\omega t)\cos\theta - \cos(\omega t)\sin\theta]$$

If the power remains constant for the time dt then

$$dW = V_p I_p[\sin^2(\omega t)\cos\theta - \sin(\omega t)\cos(\omega t)\sin\theta]$$
$$= V_p I_p[\sin^2(\omega t)\cos\theta - \frac{\sin 2\omega t}{2} - \sin\theta]$$

So the total work done or energy spent in maintaining the current over one cycle (from $t = 0$ to $t = \text{T}$) is given by

$$W = \int_0^T V_p I_p[\sin^2\omega t\cos\theta - \frac{\sin 2\omega t}{2}\sin\theta)\, dt$$

Over a complete cycle the second term

$$\int_0^T \frac{\sin 2\omega t}{2}\sin\theta\, dt = 0$$

So we can say that

$$W = V_p I_p \cos\theta \int_0^T \sin^2\omega t dt$$

To integrate $\int_0^T \sin^2\omega t dt$ we replace $\sin^2\omega t$ by $\frac{1}{2}(1 - \cos 2\omega t)$

$$\int_0^T \sin^2\omega t dt = \int_0^T \frac{1}{2}(1 - \cos 2\omega T) = \frac{T}{2} - \frac{\sin\omega T}{T}$$

Over a complete cycle the second term $\frac{\sin\omega T}{T} = 0$

Reminder about trigonometric formulae

The formula for finding the value of $\sin(A - B)$ is as follows

$\sin(A - B) = \sin A\cos B - \cos A\sin B$

We can also write

$\sin 2A = 2\sin A\cos A$

and

$\sin^2 A = \frac{1}{2}(1 - \cos 2A)$

© MoE FDR Ethiopia Not to be republished

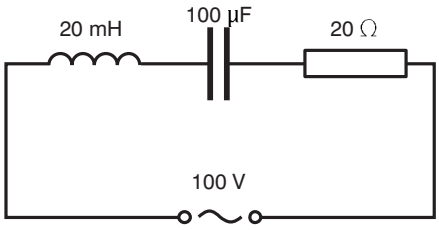


Figure 7.40

So $W = V_p I_p \cos\theta \frac{T}{2}$ and average power $= \frac{W}{T} = \frac{V_p}{\sqrt{2}} \times \frac{I_p}{\sqrt{2}} \cos\theta =$

$V_{rms} I_{rms} \cos\theta$

In a purely capacitive circuit, we know that $\theta = 90°$ and $\cos 90° = 0$ so $W = 0$.

Similarly, in a purely inductive circuit, we know that $\theta = 90°$ and $\cos 90° = 0$ so $W = 0$.

Thus the average power in a purely capacitive circuit and in a purely inductive circuit is 0 J.

The average power in an a.c. RLC circuit

We have derived an expression for the average power in an RLC circuit above

average power $= V_{rms} I_{rms} \cos\theta$

Notice that this expression is the expression we know for power for d.c. circuits multiplied by the power factor which we met in Section 7.3.

Worked example 7.19

Find the average power for the RLC circuit shown in Figure 7.41.

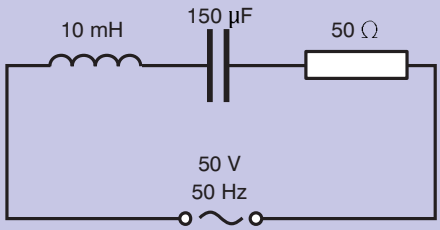


Figure 7.41

W (J)	V_{rms} (V)	I_{rms} (A)	cos θ
?	50	?	?

To find I_{rms} we need to find the impedance of the circuit.

Z (Ω)	R (Ω)	X_C (Ω)	X_L (Ω)
?	50	$\frac{1}{2\pi fC} = \frac{1}{2\pi \times 50 \times 150 \times 10^{-6}}$ $= 21.23$	$2\pi fL = 2 \times \pi \times 50 \times 10 \times 10^{-3}$ $= 3.14$

Use $Z = \sqrt{R^2 + (X_L - X_C)^2}$

$= \sqrt{50^2 + (3.14 - 21.23)^2}$

$= \sqrt{2500 + 327}$

$= \sqrt{2827}$

$= 53.2\ \Omega$

$I = \frac{V}{Z} = \frac{50}{53.2} = 0.9398$ A

To find the phase angle θ we use

$$\tan\theta = \frac{X_L - X_C}{R} = \frac{3.14 - 21.23}{50} = 424.5$$

$$\theta = \tan^{-1}(-0.3618)$$

$$= -19.89°$$

Therefore *average power* =

$$V_{rms}I_{rms}\cos\theta = 50 \times 0.9398 \times \cos(-19.89)$$

$$= 50 \times 0.9398 \times 0.9403$$

$$= 44.2 \text{ J}$$

Real, apparent and ideal power in an RLC circuit

Consider a simple a.c. circuit in which there is a supply p.d. and a linear load, such as a resistor. At every instant in such a circuit, the current and the p.d. are in phase, and only **real power** is transferred. The value of the real power is given by

real power = current^2 × resistance = supply p.d. × I_{rms}

If the load is purely reactive (a capacitor or an inductor) then the current and the supply voltage are out of phase by 90° and there is no net flow of energy to the load. The power is reactive power.

Apparent power is the vector sum of real and reactive power, as shown in Figure 7.42.

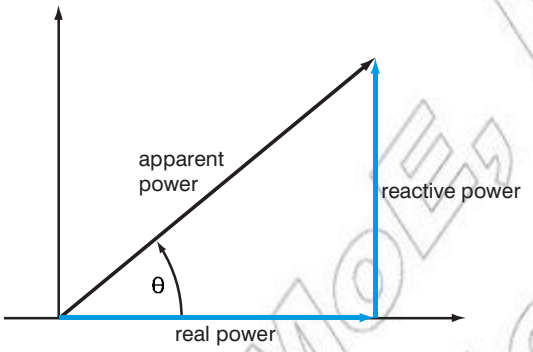


Figure 7.42

We can calculate apparent power using

apparent power = $(I_{rms})^2Z$

where Z is the impedance in the circuit.

The ideal situation (**ideal power**) is where apparent power and true power to be equal, since the difference between real power and apparent power is wasted. In this case, the power factor is 1. This can occur when either the circuit is purely resistive or where the there is no reactance. To have no reactance X_L and X_c must be equal, so that the reactance = $X_L - X_C = 0$.

KEY WORDS

real power *power transferred when the load is purely resistive*

apparent power *the vector sum of real and reactive power*

ideal power *where apparent power and true power are equal*

© MoE FDRE Ethiopia
Not to be republished

Worked example 7.20

For the circuit shown in Figure 7.43, find a) the power factor for the circuit b) the apparent power for the circuit c) the value of C needed to make the power in this circuit ideal.

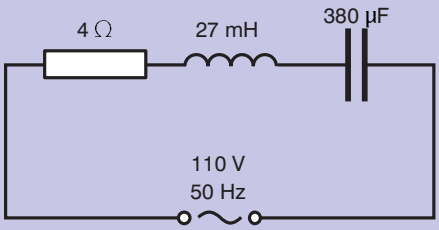


Figure 7.43

a)

cos θ	X_L (Ω)	X_C (Ω)	R (Ω)
?	2πfL	$\frac{1}{2\pi fC}$	4

$$X_L = 2\pi fL$$
$$= 2 \times \pi \times 50 \times 27 \times 10^{-3}$$
$$= 8.478\ \Omega$$

$$X_C = \frac{1}{2 \times \pi \times 50 \times 380 \times 10^{-6}}$$
$$= \frac{1}{0.11932}$$
$$= 8.38\ \Omega$$

Use $\tan^{-1}\theta = \frac{X_L - X_C}{R}$

$$= \frac{8.478 - 8.38}{4}$$
$$= 0.0245$$
$$\theta = 1.4°$$

Power factor = cos 1.4°
= 0.9997

b) Find current through the circuit.

First find impedance, Z

Use $= \sqrt{R^2 + (X_L - X_C)^2}$
$= \sqrt{4^2 + (0.098)^2}$
$= \sqrt{16 + 9.604 \times 10^{-3}}$
$= \sqrt{16.009604}$
$= 4\ \Omega$

Use $I = \frac{110}{4} = 27.5$ A

apparent power $= (I_{rms})^2 Z$
$= 27.5^2 \times 4$
$= 3025$ J

© MoE Not for

c) For ideal power need $2\pi fL = \frac{1}{2\pi fC}$

$$C = \frac{1}{(2\pi fC)^2)L}$$

$$= \frac{1}{(2 \times 3.14 \times 50)^2) \times 27 \times 10^{-3}}$$

$$= \frac{1}{98\,596 \times 27 \times 10^{-3}}$$

$$= \frac{1}{2662.092}$$

$$= 3.756 \times 10^{-4} \text{ F}$$

Summary

In this section you have learnt that:

- The average power in an a.c. capacitive or an a.c. inductive circuit is zero.
- The average power in an a.c. RLC circuit is given by the expression $V_{rms}I_{rms}\cos\theta$.
- **Real power** power transferred when the load is purely resistive.
- **Apparent power** the vector sum of real and reactive power.
- **Ideal power** where apparent power and true power are equal.

Review questions

1. Find the average power for the RLC circuit shown in Figure 7.44.
2. For the circuit shown in Figure 7.45, find a) the power factor for the circuit, b) the apparent power for the circuit, c) the value of C needed to make the power in this circuit ideal.

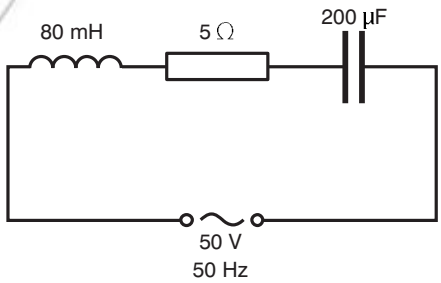


Figure 7.44

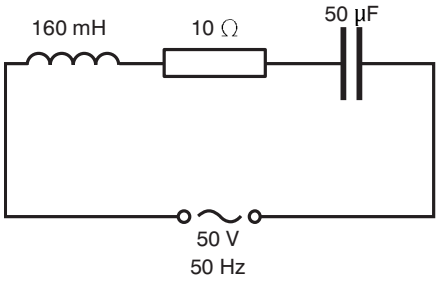


Figure 7.45

End of unit questions

1. Calculate the magnetic flux when a magnetic field of strength 5 mT passes through an area of 10 cm^2 that is at an angle of 50° to the magnetic field.
2. a) State Faraday's law of electromagnetic induction.
   b) How could you demonstrate that Lenz's law is a consequence of conservation of energy?
3. A wire 40 cm long bent into a rectangular loop of 15 cm by 5 cm is placed perpendicular to the magnetic field whose flux density is 0.8 Wb/m^2. Within 5 s the loop is changed into a 10 cm square and the flux density increases to 1.4 Wb/m^2. Find the induced e.m.f.

© Ethiopia Not to be published

4. How are eddy currents used in security scanners at airports?
5. Find the induced voltage in an inductor of 100 mH when the current is changing at a rate of 3.5 A/s.
6. a) Find the inductance of an air cored solenoid of 500 turns per unit length and area 3 cm^2. The permeability of free space is $4\pi \times 10^{-7}$ T m^2/A.
   b) Find the energy stored in the inductor in part (a) when a current of 3.5 A flows through it.
7. Find the magnetic energy density for an inductor with an air core with a magnetic field of 0.25 T. The permeability of free space is $4\pi \times 10^{-7}$ T m^2/A.
8. What values are used to give the d.c. equivalent p.d. or current when the supply is a.c.?
9. a) How does a transformer work?
   b) A transformer converts the mains supply from 220 V down to 10 V. There are 320 turns on the secondary coil. How many turns does the primary coil have?
10. What is the main reason why alternating current is used to transmit electricity?
11. a) Explain the phase difference between voltage and current in a capacitor. Why does this phase difference occur?
    b) Show the phase difference on a graph of *V* and *I* against *t*.
12. In a series RLC circuit, what determines whether the inductive or capacitive behaviour dominates?
13. Why does the amplitude of oscillations become smaller on an oscilloscope when a series RLC circuit is in resonance?
14. An RLC circuit is used in a radio to tune into an FM station broadcasting at 99.7 MHz. The resistance in the circuit is 12 Ω and the inductance is 1.4 μH. What capacitance should be used?
15. A series RLC circuit is as shown in the diagram.

    Find
    a) the resonant frequency of this circuit
    b) the amplitude of the current at the resonant frequency
    c) the amplitude of the p.d. across the inductor at resonance.
16. Find the average power for the RLC circuit shown in Figure 7.46.
17. For the circuit shown in Figure 7.47, find a) the power factor for the circuit, b) the apparent power for the circuit, c) the value of *C* needed to make the power in this circuit ideal.

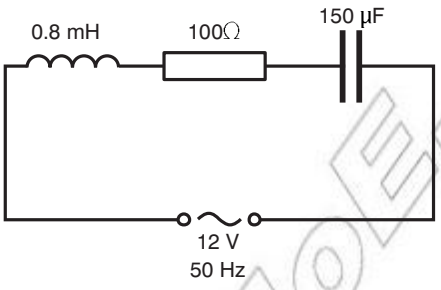


Figure 7.46

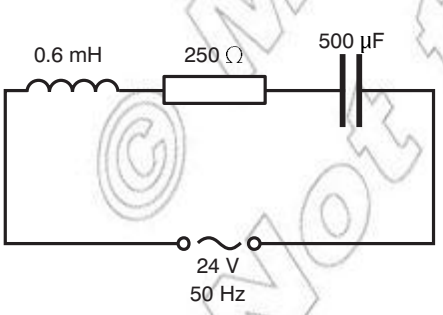


Figure 7.47

© MoE FDRE Ethiopia Not to be republished

Atomic physics

Unit 8

Contents

Section	Learning competencies
8.1 The dual nature of matter and radiation (page 312)	• Identify that black bodies absorb all electromagnetic radiation. • Describe the photoelectric effect and its characteristics. • Show understanding that matter has wave nature. • Use the de Broglie equation $\lambda = \frac{h}{p}$ to find the wavelength of a matter particle. • State Heisenberg's uncertainty principle. • Use the uncertainly principle to relate the uncertainties in position and momentum. • Find the uncertainty in position from the uncertainty in momentum.
8.2 Atoms and nuclei (page 322)	• Describe Rutherford's model of the atom. • Describe Bohr's model of the atom. • Show understanding that electrons can only exist at specific energy states, and will not be found with energies between those levels. • Compute the change in energy of an atom using the relation $\Delta E = E_f - E_i$. • Represent diagrammatically the structure of simple atoms. • Use the relationship $A = Z + N$ to explain what is meant by the term isotope. • Compare the charge and mass of the electron with the charge and mass of the proton. • Identify nuclear force is a very strong force that holds particles in a nucleus together. • State some important properties of the strong force. • Show radius and mass number are related mathematically $R = (1.2 \times 10^{-15}\ \text{m})A^{1/3}$. • State the approximate size of an atom. • State nuclear properties. • Explain how nuclear stability is determined by binding energy per nucleon. • Define the term binding energy. • Compare graphs of stable and unstable nuclei. • Interpret graphs of binding energy per nucleon versus mass number. • Associate radioactivity with nuclear instability. • Define the term nuclear fission. • Define the term nuclear fusion. • Distinguish between fission and fusion. • Show understanding that radioactivity emission occurs randomly over space. • Identify that the decay process is independent of conditions outside the nucleus.

Contents

Section	Learning competencies
	• Identify the nature of the three types of emissions from radioactive substances. • Distinguish between the three kinds of emissions in terms of their nature, relative ionising effect, relative penetrating power. • Describe the need for safety measures in handling and using radioisotopes. • Describe experiments to compare the range of alpha, beta and gamma radiation in various media. • Predict the effect of magnetic and electric fields on the motion of alpha, beta and gamma rays. • Name the common detectors for α–particles, β–particles and γ–rays. • Associate the release of energy in a nuclear reaction with a change in mass. • Apply quantitatively the laws of conservation of mass and energy, using Einstein's mass–energy equation. • Represent and interpret nuclear reactions of the form ${}^{14}_{6}C \rightarrow {}^{14}_{6}N + {}^{0}_{-1}e$ (beta). • Represent nuclear reactions in the form of equations. • Define the term half-life. • Work through simple problems on half-life. • Use graphs of random decay to show that such processes have a constant half-life. • State the uses of radioactive isotopes. • Discuss problems posed by nuclear waste.

8.1 Dual nature of matter and radiation

By the end of this section you should be able to:

- Identify that black bodies absorb all electromagnetic radiation.
- Describe the photoelectric effect and its characteristics.
- Show understanding that matter has wave nature.
- Use the de Broglie equation $\lambda = \frac{h}{p}$ to find the wavelength of a matter particle.
- State Heisenberg's uncertainty principle.
- Use the uncertainly principle to relate the uncertainties in position and momentum.
- Find the uncertainty in position from the uncertainty in momentum.

© MoE Not for sale

Black bodies

Electromagnetic radiation is given off across a wide range of wavelengths (the electromagnetic spectrum) and is emitted by all objects. Our eyes can only see electromagnetic radiation in the visible part of the spectrum, and radiation from hot objects such as glowing coals on a fire.

A black body is an object that is a perfect radiator of electromagnetic energy – it will radiate energy over the entire electromagnetic spectrum, as shown in Figure 8.1.

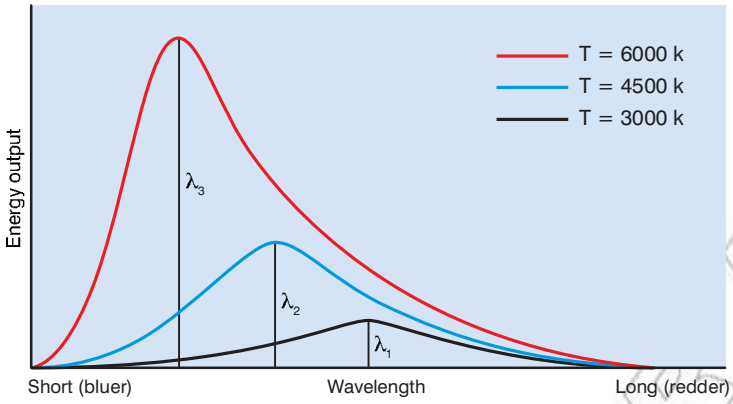


Figure 8.1

A **black body** will also absorb energy over the entire electromagnetic spectrum.

Worked example 8.1

In a cold country on a sunny day, a lady decides to try and melt snow by sprinkling a thin layer of soot over it. Explain her thinking.

The thin layer of soot will cover the snow and make it appear to be a black body. This black body will then absorb all the radiation emitted by the Sun and this energy will melt the snow.

The photoelectric effect and its characteristics

When light, particularly ultraviolet light, is shone on a clean metal surface, the surface will emit electrons. These electrons are known as photoelectrons, and the emission of photoelectrons is called the **photoelectric effect**.

The energy of the photoelectrons is proportional to the frequency of the radiation from the ultraviolet lamp. We can use the equation

$E = hf$

where E is the energy of the photoelectron, f is the frequency of the radiation and h is a constant known as the Planck constant, which has been shown to be 6.63×10^{-34} J s to 3 significant figures, to find the energy of the photoelectrons.

KEY WORDS

black body *an object that is a perfect radiator and absorber of electromagnetic energy*

photoelectric effect *emission of photoelectrons from metal surface when light is shone on the surface*

DID YOU KNOW?

The German physicist who first noticed the photoelectric effect, Heinrich Hertz, showed an aptitude for languages while studying in Hamburg, learning Arabic and Sanskrit. He supplied the first experimental evidence for the existence of radio waves, generating them by means of an electric spark.

© Ethiopia EDK Not to be republished

For radiation in the electromagnetic spectrum, we know that the speed of radiation, $c = 3 \times 10^8$ m/s and that $c = f\lambda$ where f is the frequency and λ is the wavelength of the radiation. So our equation for the energy of the photoelectrons can also be written

$$E = \frac{hc}{\lambda}$$

Activity 8.1: Demonstrating the photoelectric effect

You can demonstrate the photoelectric effect using the apparatus shown in Figure 8.2.

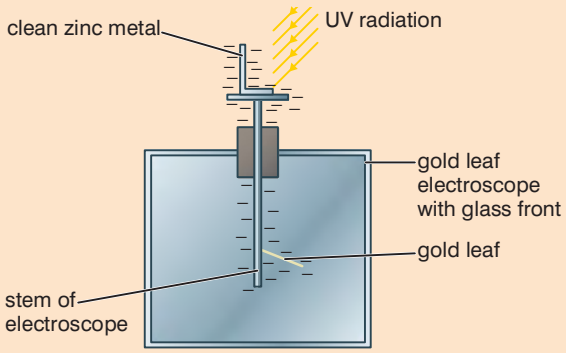


Figure 8.2

Place a piece of clean zinc metal on the electroscope. The metal will become negatively charged. The stem of the electroscope will also become negatively charged and the negatively charged gold leaf will be repelled from the stem.

Now bring an ultraviolet lamp near to the zinc. The gold leaf will start to fall back towards the stem. Remove the lamp. The leaf will stop falling. This means that the ultraviolet light from the lamp is causing electrons to be emitted from the zinc – the photoelectric effect.

Worked example 8.2

Ultraviolet radiation of wavelength 4×10^{-8} m falls on a sample of zinc. Photoelectrons are emitted. Calculate the energy of these electrons. Take the Planck constant to be 6.63×10^{-34} J s and the speed of light to be 3×10^8 m/s.

E (J)	h (J s)	c (m/s)	λ (m)
?	6.63×10^{-34}	3×108	4×10^{-8}

Use $E = \frac{hc}{\lambda}$

$$= \frac{6.63 \times 10^{-34} \times 3 \times 10^8}{4 \times 10^{-8}}$$

$$= 4.97 \times 10^{-18} \text{ J}$$

© MoE Not for sale

Electronvolts

The energy of photoelectrons is very low, as shown by worked example 8.2. We use a unit called the **electronvolt** (eV) to describe photoelectron energy.

One electronvolt is the energy change of an electron when it moves through a potential difference of one volt.

Since we know that one volt is one joule per coulomb and the charge on an electron is 1.6×10^{-19} C, we get

$1 \text{ eV} = 1 \text{ J/C} \times 1.6 \times 10^{-19} \text{ C} = 1.6 \times 10^{-19} \text{ J}$

So we can express the answer to worked example 8.2 as

4.97×10^{-18} J or $\frac{4.97 \times 10^{-18}}{1.6 \times 10^{-19}} = 31.1$ eV

Activity 8.2: Investigating the energy of the photoelectrons

You could use the circuit shown in Figure 8.3 to measure the energy of the photoelectrons.

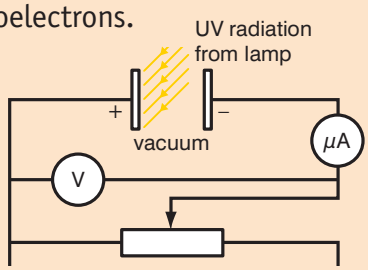


Figure 8.3

Apply a p.d. between the source of the electrons and the metal plate. The photoelectrons have a negative charge, so they are repelled by the negative plate and attracted to the positive one. If, however, the electrons have sufficient kinetic energy they are able to overcome the repulsion of the negative plate and reach it and you can record a current (a photocurrent).

Vary the potential between the plates from negative values to positive values and observe how the photocurrent varies. Record your results and draw a graph of photocurrent against p.d. between the plates.

From Activity 8.2, you should obtain a graph like the one shown in Figure 8.4.

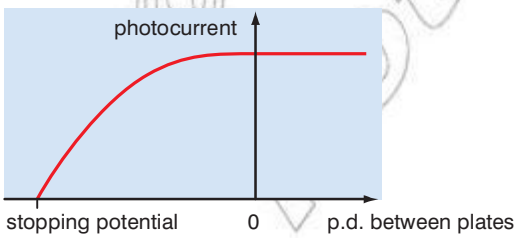


Figure 8.4

The value of the stopping potential marked on the graph can be equated to the maximum energy of the photoelectrons. The

KEY WORDS

electronvolt *the energy change of an electron when it moves through a potential difference of one volt*
$1 \text{ eV} = 1.6 \times 10^{-19}$ J

Think about this...

If the light shining on the metal were more intense, would more or fewer photoelectrons be emitted?

If the light shining on the metal were more intense, what would happen to the maximum kinetic energy of the photoelectrons?

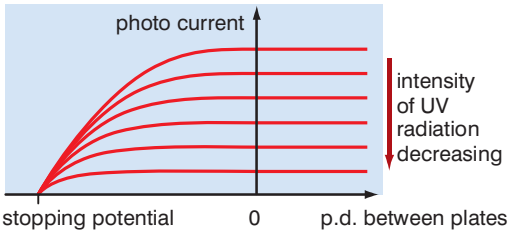


Figure 8.5

electronvolt is useful here, since if the stopping potential is, for example, 1.85 V, then we can say that the maximum energy for the photoelectrons is 1.85 eV.

You might think that if the light shining on a metal to produce photoelectrons were more intense, more electrons would be emitted, and vice versa. This is indeed the case. You might also think that the maximum kinetic energy of the photoelectrons would also be related to the intensity of the radiation that caused them to be emitted. However, experimentally we find that the maximum kinetic energy of the photoelectrons remains constant, as shown in Figure 8.5.

This result was unexpected and it led Einstein to suggest that some of the ultraviolet radiation hitting the surface of the metal was used to release an electron from the surface atom of the metal, and any remaining energy became the kinetic energy of the electron. Using the law of conservation of energy gives the equation

photon energy = energy to release electron + kinetic energy of the electron

$$hf = \varphi + \left(\frac{1}{2}mv^2\right)_{max}$$

The minimum energy required to release the electron, φ, is called the **work function** for the metal.

This equation is called the Einstein photoelectric equation.

It can be rearranged to give

$$KE_{max} = hf - \varphi$$

KEY WORDS

work function *minimum energy required to release the electron from surface of metal*

Worked example 8.3

An ultraviolet lamp is used to illuminate a clean zinc surface and photoelectrons are emitted. The stopping potential of the photoelectrons is –1.92 V. The work function energy of zinc is 4.24 eV. Calculate the wavelength of the UV radiation. Take the value of the Planck constant to be 6.63×10^{-34} J s.

photon energy (eV)	φ (eV)	KE_{max} (eV)
?	4.24	1.92

Use

$$hf = \varphi + \left(\frac{1}{2}mv^2\right)_{max}$$

$$= 4.24 + 1.92$$

$$= 6.16 \text{ eV}$$

$$6.16 \text{ eV} = 6.16 \times 1.6 \times 10^{-19} \text{ J}$$

$$= 9.856 \times 10^{-19} \text{ J}$$

$$E = hf = \frac{hc}{\lambda}$$

$$\lambda = \frac{hc}{E} = \frac{6.63 \times 10^{-34} \times 3 \times 10^{8}}{9.856 \times 10^{-19}}$$

$$= 2.02 \times 10^{-7} \text{ m}$$

$$= 202 \text{ nm}$$

© Molteno Institute for Physics Not to be republished

The wave nature of matter

The photoelectric effect provided evidence that electromagnetic waves could be considered as a stream of particles called photons. This evidence was reinforced when, in 1922, Arthur Compton discovered that X-rays were colliding with electrons and behaving as particles.

In 1924, Louis de Broglie suggested that an atom's electrons, protons and neutrons possess both particle and wave properties. This is known as the **wave–particle duality of matter**.

To prove that particles can also act as waves, you need to show particles exhibiting a behaviour that is also demonstrated by waves, such as interference or diffraction. It is possible to diffract electrons using the apparatus shown in Figure 8.6.

Electrons from an electron gun are accelerated through a vacuum towards a layer of graphite. (The atomic spacing in graphite is the right order of magnitude for electrons to be diffracted.) A circular diffraction pattern such as the one shown in Figure 8.7 is obtained.

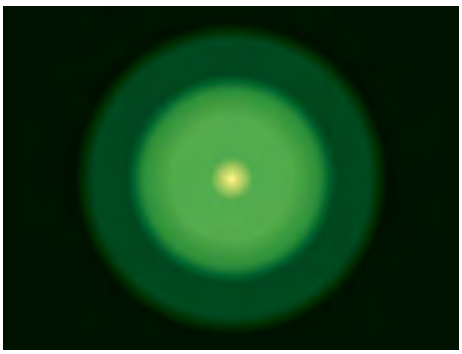


***Figure 8.7** An electron diffraction pattern*

You can calculate the wavelength of the electrons by measuring the ring diameters in the diffraction pattern. It can be shown that the wavelength, λ, is inversely proportional to the speed, v, of the electron. Further evidence gives the equation

$$\lambda = \frac{h}{mv} = \frac{h}{p}$$

where m is the mass of the electron, h is the Planck constant and p is the momentum (mv) of the electron. This equation is known as the de Broglie equation.

KEY WORDS

wave–particle duality of matter *theory that matter possesses both wave and particle properties, depending on circumstances*

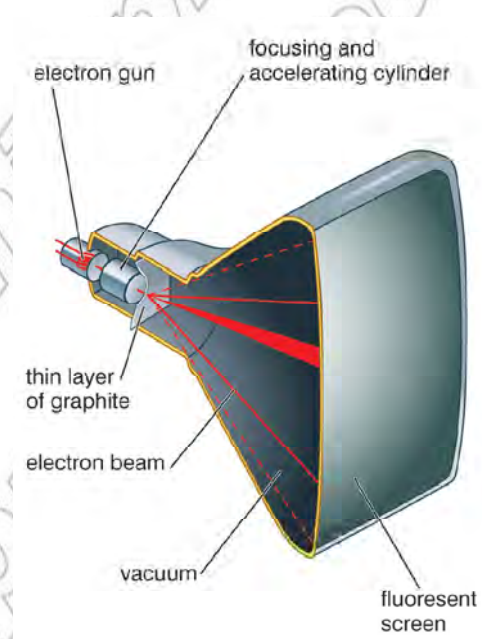


Figure 8.6

DID YOU KNOW?

Electron diffraction can be used to determine atomic spacing. High-speed electrons can be used to measure the diameter of a nucleus.

© Model FDR Exhibition Not to be reproduced

Worked example 8.4

Electrons are accelerated through a potential difference of 4000 V before striking a layer of graphite and being diffracted. The mass of an electron is 9.11×10^{-31} kg. Planck's constant is 6.63×10^{-34} J s.

Calculate

a) the speed of the electrons when they hit the layer of graphite

b) the momentum of the electrons

c) the wavelength of the electrons.

a) 4000 V is 4000 J/C

An electron has a charge of 1.6×10^{-19} C so its kinetic energy will be $4000 \times 1.6 \times 10^{-19}$ J = 6.4×10^{-16} J

KE (J)	m (kg)	v (m/s)
6.4×10^{-16}	9.11×10^{-31}	?

Use $KE = \frac{1}{2}mv^2$

$$v = \sqrt{\frac{2 \times KE}{m}}$$

$$= \sqrt{\frac{2 \times 6.4 \times 10^{-16}}{9.11 \times 10^{-31}}}$$

$$= \sqrt{\frac{1.28 \times 10^{-15}}{9.11 \times 10^{-31}}}$$

$$= 3.7 \times 10^{7} \text{ m/s}$$

b)

p (kg m/s)	m (kg)	v (m/s)
?	9.11×10^{-31}	3.7×10^{7}

Use $p = mv$

$$= 9.11 \times 10^{-31} \times 3.7 \times 10^{7}$$

$$= 3.37 \times 10^{-23} \text{ kg m/s}$$

c)

λ (m)	h (J s)	p (kg m/s)
?	6.63×10^{-34}	3.37×10^{-23}

Use $\lambda = \frac{h}{p}$

$$= \frac{6.63 \times 10^{-34}}{3.37 \times 10^{-23}}$$

$$= 1.97 \times 10^{-11} \text{ m}$$

© MoE - Not for sale

Activity 8.4: The wave and particle nature of photons and electrons

In a small group, discuss the wave and particle nature of photons and electrons. Discuss an experiment where a photon behaves like a particle and an experiment where it behaves like a wave.

Heisenberg's uncertainty principle

We know that the wavelength and momentum of a particle such as an electron are related by the de Broglie equation, $\lambda = \frac{h}{p}$.

However, Heisenberg's uncertainty principle states that it is impossible to know both the exact position and the exact velocity (and therefore momentum) of a particle at the same time.

In order to observe any particle, smaller particles must be reflected off it. For example, to find the position and momentum of an electron, at least one photon of light must be used. The photon must hit the electron and then be reflected back to the measuring device. If objects are large, such as sand grains or buses, the percentage uncertainty in the measurements of position and momentum are insignificant. However, for subatomic particles, which are much smaller, the percentage uncertainty in the position and momentum measurements is far larger, and finding momentum and position becomes more difficult.

An electron has momentum p = mass × velocity

When photons are moving, they have an apparent mass due to their kinetic energy. When one of the photons bounces off the electron, the momentum of the electron will be changed (in the same way as the momentum of a billiard ball is changed when it collides with another ball or the sides of the table). This change in momentum ($\Delta mv = \Delta p$) is uncertain and will be of the same order of magnitude as the photon's momentum, so we can say that $\Delta p \approx \frac{h}{\lambda}$.

The photon of light cannot measure the electron's position (x) with perfect accuracy. To increase the accuracy of the position measurement, a photon with a smaller wavelength must be used. To see why this is the case, consider Figure 8.8, which shows two photons with different wavelengths, with the position of the electron marked.

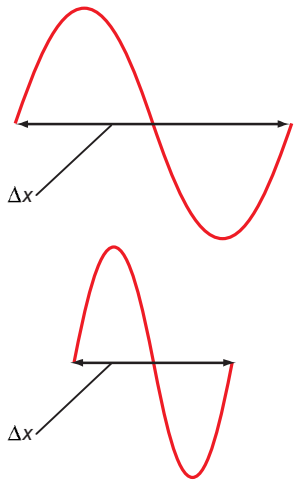


Figure 8.8

You can see that the position of the electron, which must be somewhere in the distance Δx in each case, can be more accurately determined with photon b) than with photon a).

However, decreasing the wavelength increases the frequency and thus the energy of the photon. Therefore, when the photon collides with the electron, it will change the momentum of the electron by a larger amount, so if the position is more accurately determined the momentum is less accurately determined. The opposite is also true – if the momentum is more accurately determined then the position must be less accurately determined.

© Ministry of Education Ethiopia Not to be republished

An estimate of the position of the particle is at least one photon wavelength from the original position.

The uncertainty in position, $\Delta x \geq \lambda$.

We know that $\Delta p \approx \frac{h}{\lambda}$

We can combine the equations to give

$\Delta p \Delta x \geq \frac{h}{\lambda} \times \lambda \geq h$

Worked example 8.5

Light of wavelength 3×10^{-6} m is used to measure the position of an electron of wavelength 1.97×10^{-11} m. Find the uncertainty in the position of the electron. Planck's constant is 6.63×10^{-34} J s.

Δp (kg m/s)	h (J s)	λ (m)
?	6.63×10^{-34}	1.97×10^{-11}

Use $\Delta p \approx \frac{h}{\lambda}$

$\Delta p \approx \frac{6.63 \times 10^{-34}}{1.97 \times 10^{-11}} \approx 3.37 \times 10^{-23}$ kg m/s

Δx (m)	h (J s)	Δp (kg m/s)
?	6.63×10^{-34}	3.37×10^{-23}

Use $\Delta p \Delta x \geq h$

$\Delta x \geq \frac{h}{\Delta p}$

$\geq \frac{6.63 \times 10^{-34}}{3.37 \times 10^{-23}}$

$\geq 1.97 \times 10^{-11}$ m

© MoE Not to be republished

Summary

In this section you have learnt that:

- Black bodies absorb all electromagnetic radiation.
- The **photoelectric effect** is the emission of photoelectrons from metal surface when light is shone on the surface.
- The minimum energy required to release the electron, φ, is called the **work function** for the metal.
- $KE_{max} = hf - \varphi$
- Matter has a wave nature as well as a particle nature.
- The de Broglie equation $\lambda = \frac{h}{p}$ is used to find the wavelength of a matter particle.
- Heisenberg's uncertainty principle is that it is impossible to know both the exact position and the exact velocity (and therefore momentum) of a particle at the same time.
- The uncertainties in position and momentum are related by the equation $\Delta p \Delta x \geq h$

Review questions

1. What is a black body?
2. Ultraviolet radiation of wavelength 200 nm falls on a sample of magnesium. Photoelectrons are emitted. Calculate the energy of these electrons. Take the Planck constant to be 6.63×10^{-34} J s and the speed of light to be 3×10^{8} m/s.
3. An ultraviolet lamp is used to illuminate a clean lithium surface and photoelectrons are emitted. The stopping potential of the photoelectrons is –1.92 V. The work function energy of lithium is 2.93 eV. Calculate the wavelength of the UV radiation. Take the value of the Planck constant to be 6.63×10^{-34} J s.
4. Electrons are accelerated through a potential difference of 5000 V before striking a layer of graphite and being diffracted. The mass of an electron is 9.11×10^{-31} kg. Planck's constant is 6.63×10^{-34} J s.

   Calculate

   a) the speed of the electrons when they hit the layer of graphite

   b) the momentum of the electrons

   c) the wavelength of the electrons.
5. State Heisenberg's uncertainty principle.
6. Light of wavelength 3×10^{-6} m is used to measure the position of an electron of wavelength 1.73×10^{-11} m. Find the uncertainty in the position of the electron. Planck's constant is 6.63×10^{-34} J s.

© Ministry of Education Not to be republished

8.2 Atoms and nuclei

By the end of this section you should be able to:

- Describe Rutherford's model of the atom.
- Describe Bohr's model of the atom.
- Show understanding that electrons can only exist at specific energy states, and will not be found with energies between those levels.
- Compute the change in energy of an atom using the relation $\Delta E = E_f - E_i$.
- Represent diagrammatically the structure of simple atoms.
- Use the relationship $A = Z + N$ to explain what is meant by the term isotope.
- Compare the charge and mass of the electron with the charge and mass of the proton.
- Identify nuclear force is a very strong force that holds particles in a nucleus together.
- State some important properties of the strong force.
- Show that radius and mass number are related mathematically $R = (1.2 \times 10^{-15}\ \text{m})A^{1/3}$.
- State the approximate size of an atom.
- State nuclear properties.
- Explain how nuclear stability is determined by binding energy per nucleon.
- Define the term binding energy.
- Compare graphs of stable and unstable nuclei.
- Interpret graphs of binding energy per nucleon versus mass number.
- Associate radioactivity with nuclear instability.
- Define the term nuclear fission.
- Define the term nuclear fusion.
- Distinguish between fission and fusion.
- Show understanding that radioactivity emission occurs randomly over space.
- Identify that the decay process is independent of conditions outside the nucleus.
- Identify the nature of the three types of emissions from radioactive substances.

© MoE Not

- Distinguish between the three kinds of emissions in terms of their nature, relative ionising effect, relative penetrating power.
- Describe the need for safety measures in handling and using radioisotopes.
- Describe experiments to compare the range of alpha, beta and gamma radiation in various media.
- Predict the effect of magnetic and electric fields on the motion of alpha, beta and gamma rays.
- Name the common detectors for α–particles, β–particles and γ–rays.
- Associate the release of energy in a nuclear reaction with a change in mass.
- Apply quantitatively the laws of conservation of mass and energy, using Einstein's mass–energy equation.
- Represent and interpret nuclear reactions of the form ${}^{14}_{6}C \rightarrow {}^{14}_{6}N + {}^{0}_{-1}e$ (beta).
- Represent nuclear reactions in the form of equations.
- Define the term half-life.
- Work through simple problems on half-life.
- Use graphs of random decay to show that such processes have a constant half-life.
- State the uses of radioactive isotopes.
- Discuss problems posed by nuclear waste.

Rutherford's model of the atom

Before the 20th century, there had been various models of the atom. In 1906, Thomson discovered that electrons could be removed from the atom and proposed that the main part of the atom was positively charged and had negatively charged electrons scattered through it.

Between 1909 and 1911, two of Lord Rutherford's students, Geiger and Marsden, aimed charged particles at an extremely thin piece of gold foil. They used apparatus similar to that shown in Figure 8.9.

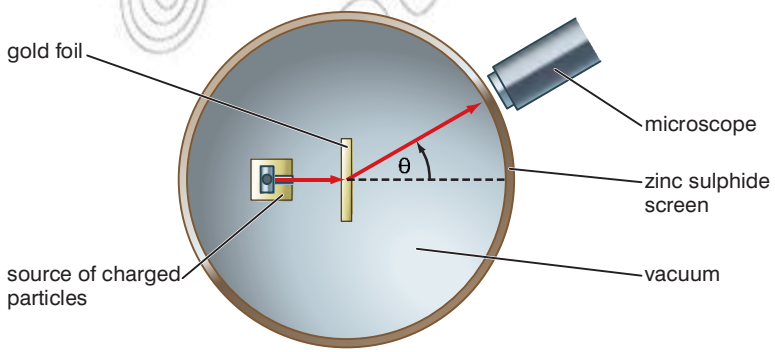
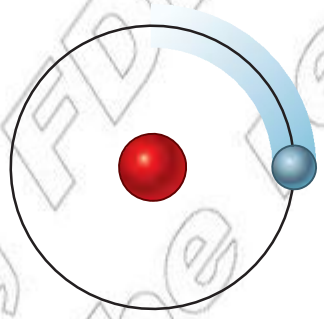


Figure 8.9

They expected that all the charged particles would pass through the foil, perhaps with a little deviation. This was the case with most of the particles. However, some of them were deflected at quite large angles – some were even sent back the way they had come. This result was very surprising if Thomson's model of the atom was correct. Rutherford repeated the experiment many times, with the same result. He was forced to conclude that, in order to explain the results, which are shown in the table, the model of the atom needed to be changed.

Angle of deflection (°)	Evidence	Conclusion
0–10	Most charged particles pass through with little deviation	Most of the atom is empty space
10–90	Some charged particles deflected through large angle concentrated in one place	All the atom's positive charge
90–180	A few charged particles sent back to source	Most of the mass, and all positive charge, is in tiny, central nucleus

Rutherford's model of the atom is as shown in Figure 8.10.



Rutherford's nuclear model of the atom: All the positive charge and most of the mass is concentrated in a tiny central nucleus. Most of the atom is empty space, and electrons orbit at the edge.

Figure 8.10

DID YOU KNOW?

Lord Rutherford supervised many Nobel Prize winners: Chadwick for discovering the neutron (in 1932), Cockcroft and Walton for an experiment which was to be known as splitting the atom using a particle accelerator, and Appleton for demonstrating the existence of the ionosphere.

Bohr model of the atom

Niels Bohr proposed that specific energy levels existed within an atom's structure and that electrons move in circular orbits around the nucleus. In his model, electrons that are closer to the nucleus have a lower energy state than those that are further away. It is possible for electrons to exist in different states, but as they move from a higher energy level, E_1, to a lower energy level, E_2, they emit radiation. The energy of a quantum of this radiation is given by the equation

$$hf = E_1 - E_2$$

Electrons can only exist in specific energy states and will not be found with energies between these states.

Bohr's model of the atom allows line spectra to be explained. Only certain frequencies are present in line spectra and each element has a unique pattern as shown in Figure 8.11. Such spectra are produced by hot gases, where atoms are far apart.

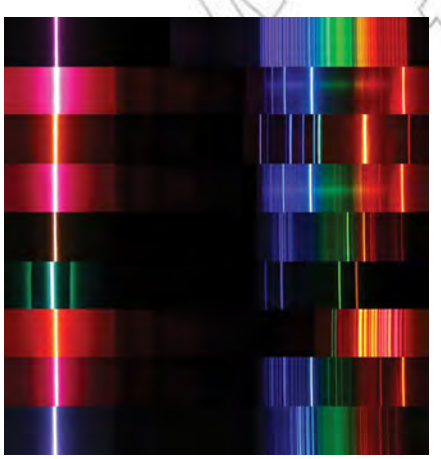


Figure 8.11

Worked example 8.6

Two lines in the spectrum of a sodium lamp have wavelengths of 589 nm and 589.6 nm.

a) What are the frequencies of these wavelengths?

b) What energy changes do these transitions correspond to in electronvolts?

Take $c = 3 \times 10^8$ m/s, $h = 6.63 \times 10^{-34}$ J s and 1 eV = 1.6×10^{-19} J.

a)

c (m/s)	f (Hz)	λ (m)
3×10^8	?	589×10^{-9}
3×10^8	?	589.6×10^{-9}

Use $f =$

For $\lambda = 589 \times 10^{-9}$ m

$$f = \frac{3 \times 10^8}{589 \times 10^{-9}}$$

$$= 5.09 \times 10^{14} \text{ Hz}$$

For $\lambda = 589.6 \times 10^{-9}$

$$f = \frac{3 \times 10^8}{589.6 \times 10^{-9}}$$

$$= 5.088 \times 10^{14} \text{ Hz}$$

b) For $\lambda = 589 \times 10^{-9}$ m

$E_1 - E_2$ (eV)	h (J s)	f (Hz)
?	6.63×10^{-34}	5.09×10^{14} Hz

Use $E_1 - E_2 = hf$

$$= 6.63 \times 10^{-34} \times 5.09 \times 10^{14}$$

$$= 3.375 \times 10^{-19} \text{ J}$$

$$= \frac{3.375 \times 10^{-19}}{1.6 \times 10^{-9}} \text{ eV}$$

$$= 2.109 \text{ eV}$$

For $\lambda = 589.6 \times 10^{-9}$ m

$E_1 - E_2$ (eV)	h (J s)	f (Hz)
?	6.63×10^{-34}	5.088×10^{14} Hz

Use $E_1 - E_2 = hf$

$$= 6.63 \times 10^{-34} \times 5.088 \times 10^{14}$$

$$= 3.373 \times 10^{-19} \text{ J}$$

$$= \frac{3.373 \times 10^{-19}}{1.6 \times 10^{-19}} \text{ eV}$$

$$= 2.108 \text{ eV}$$

Activity 8.4: To demonstrate a simple absorption spectrum

Crush some leaves in an alcohol solution. Use a diffraction grating spectrometer or a prism to show that the solution will absorb at both ends of the spectrum.

The model of the atom used now

The model of the atom that we use now was proposed by Heisenberg in the 1920s. We know that Heisenberg's uncertainty principle means that we cannot know the precise position and velocity of a particle such as an electron at any given moment. In Heisenberg's model, there is a central nucleus around which there are regions in which there is a high probability of finding an electron, and the shape of these 'probability clouds' represent what we refer to as the electron 'orbitals' (see Figure 8.12).

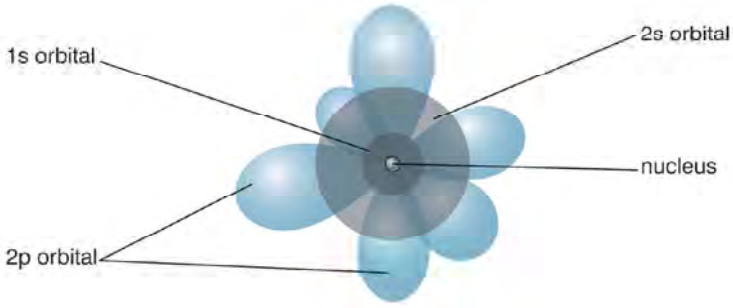


Figure 8.12

You may have met a simplified model in chemistry, such as the one shown in Figure 8.13. This is called a dot and cross diagram – the crosses represent electrons and circles show the shells. The nucleus is shown as a circle in the middle.

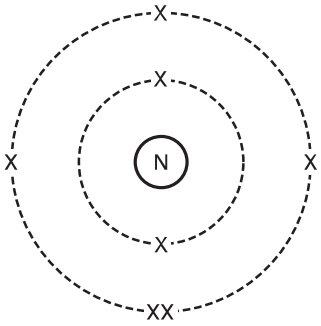


Figure 8.13

Activity 8.5: Representing the structure of a simple atom diagrammatically

The element carbon has six electrons, two in an inner shell and four in an outer shell. Use this information to represent the structure of carbon using a dot and cross diagram like the one in Figure 8.13.

Dot and cross diagrams represent the distribution of electrons in an atom but they do not give any detail about what is in the nucleus. We now know that there are two types of particles in most atomic nuclei: protons (which are positively charged) and neutrons which do not have any charge. The collective name for these particles is nucleons. The number of protons in the nucleus determines what element the atom is: carbon atoms have a different number of protons from nitrogen atoms, for example. The number of protons in an uncharged atom is the same as the number of electrons. Check that you understand why we can say that a nitrogen nucleus has 7 protons and a carbon nucleus has 6 protons! The elements in the periodic table are listed in order of proton number, which is also called the atomic number and is often given the symbol Z. The number of neutrons must be at least as great as the number of protons, but some elements have more neutrons than protons. There are some elements that have different forms of atom, which have the same number of protons (so they are the same element) but different numbers of neutrons. For example, carbon can have

© MoE. Not to be republished

atoms with 6 neutrons, atoms with 7 neutrons and atoms with 8 neutrons (but all these have 6 protons). These different atoms are called **isotopes** of carbon.

We can use a chemical shorthand to describe atoms of elements. We combine the chemical symbol for the element with the atomic number, Z, as a subscript, and the mass number, A (which is the total number of protons, Z, plus the total number of neutrons, N, so $A = Z + N$) as a superscript. For example, we can write the isotopes of carbon mentioned above as shown in Figure 8.14.

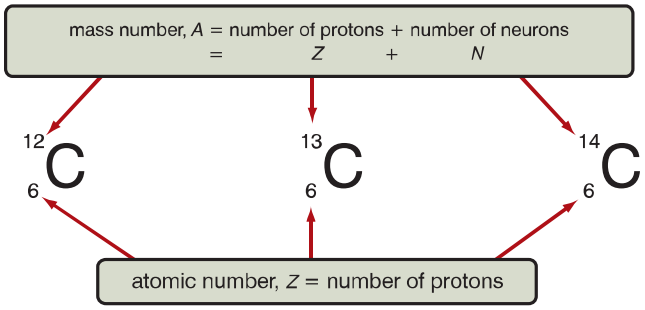


Figure 8.14

KEY WORDS

isotope *atoms of the same element that have different numbers of neutrons*

Atomic mass unit

A precise definition for the atomic mass unit is that it is one twelfth of the mass of an isolated atom of carbon-12 (^{12}C) at rest and in its ground state. A simpler definition is that it is the mass of a proton or a neutron. It is equivalent to 1.6×10^{-27} kg.

The charge and mass of electrons and protons

We know that the charge on a proton is positive and the charge on an electron is negative. The mass of a proton is 1 atomic mass unit (1.6×10^{-27} kg). The mass of an electron is taken to be $\frac{1}{1836}$ that of a proton.

Activity 8.6: Isotopes of chlorine

Chlorine has two isotopes, chlorine-35 and chlorine-37.

a) Write down the shorthand for each of these isotopes. The chemical symbol for chlorine is Cl and it has 17 protons.

b) How many neutrons are there in the nucleus of the atom of each of these isotopes?

The strong nuclear force

The protons and neutrons in the nucleus of an atom are held together by one of the four basic forces in nature, the strong nuclear force (the others are gravity, the electromagnetic force and the weak nuclear force). It is the strongest of the four forces but it has the shortest range, so particles have to be extremely close together before its effects are felt. It is strong enough to overcome the repulsion between the positive charges on protons. It is created between nucleons by the exchange of particles called mesons, as shown in Figure 8.15. An analogy for this exchange is a tennis ball being constantly hit backwards and forwards between two players. As long as this exchange can happen, the strong force can hold the nucleons together.

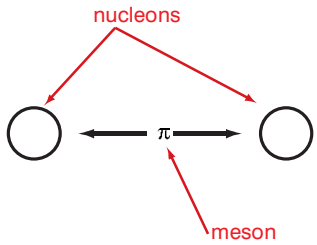


Figure 8.15

The nucleons must be extremely close for this exchange to happen – the distance must be about the diameter of a proton or a neutron. If the nucleons are unable to get this close, the strong force is too weak to make them stick together and other competing forces (usually the electromagnetic force) will influence the particles

to move apart. In Figure 8.16, the dotted line represents any electrostatic repulsion that might be present because of the charges of the nucleons that are involved. A particle must be able to cross this barrier in order for the strong force to 'glue' the particles together.

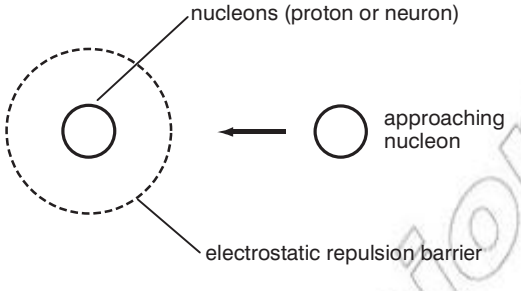


Figure 8.16

If the approaching nucleon in Figure 8.16 is a proton or another nucleus, the closer they get, the more they feel the repulsion from the other proton or nucleus. In order to get two protons or nuclei close enough to begin exchanging mesons, they must either be moving extremely fast (which means the temperature must be extremely high) or they must be under immense pressure so that they are forced to be close enough together to allow the exchange of mesons needed for the strong force. The temperature and pressure could both be extremely high, which would also allow the strong force to operate.

Neutrons in the nucleus help to reduce the repulsion between protons in the nucleus. They have no charge and so do not add to the repulsion already present, but they help separate the protons so that they do not feel so much repulsion from other protons and the neutrons also add to the strong nuclear force since they participate in meson exchange. These factors, together with the fact that protons are tightly packed in the nucleus so that they can exchange mesons, create enough strong force for the protons to overcome the repulsion between them and allow the nucleons to stay bound together.

There is evidence that the strong force is the same between any pair of nucleons. Electron scattering experiments suggest that nuclei are roughly spherical and appear to have a constant density. The data are summarised in the Fermi model as

$$r = 1.2 \times 10^{-15} \times A^{\frac{1}{3}} \text{ m}$$

where r is the radius of the nucleus and A is the mass number for the atom. Experimental results show that the radii of atoms vary from 35×10^{-12} m for hydrogen atoms where A is 1, to 175×10^{-12} m for americium atoms where A is 95.

© MoE-FDRE Ethiopia Not to be republished

Worked example 8.7

The mass number of oxygen is 16. Calculate the radius of an oxygen nucleus.

r (m)	A
?	16

Use $r = 1.2 \times 10^{-15} \times 16^{\frac{1}{3}}$ m

$= 1.2 \times 10^{-15} \times 16^{\frac{1}{3}}$

$= 1.2 \times 10^{-15} \times 2.52$

$= 3.024 \times 10^{-15}$ m

Nuclear properties

Nuclei of atoms can be ordered according to atomic number and number of nucleons. When this is done, the following properties are observed.

- For the lighter nuclei, if we look at the most common isotope, N is approximately equal to Z.
- As we get to heavier nuclei, past $Z = 20$, we begin to see N considerably greater than Z. As nuclei get heavier this becomes more apparent.
- Bismuth is the heaviest stable nucleus. Heavier nuclei exist but they are all unstable – they undergo certain spontaneous changes which we observe as radioactivity (see page 330). Nuclei from $Z = 84$ (polonium) to 92 (uranium) are found in nature (on Earth) and all their isotopes are radioactive.
- Nuclei heavier than uranium exist but they are all artificial – they have been created by scientists in laboratories. The heaviest known nucleus has $Z = 118$. It was produced in 2006.

Nuclear stability

We know, from page 327, that an element may have several isotopes (nuclei with the same number of protons but different numbers of neutrons). Isotopes are known collectively as nuclides. About 256 (76%) of the nuclides that occur naturally on Earth have not been observed to decay and are therefore referred to as 'stable isotopes'. For 80 of the chemical elements, there is at least one stable isotope. The average number of stable isotopes per element among those that have stable isotopes is 3.1. Twenty-seven elements have only a single stable isotope, while the largest number of stable isotopes observed for any element is ten, for the element tin. Elements 43, 61, and all elements numbered 83 or higher have no stable isotopes.

The stability of isotopes is affected by the ratio of protons to neutrons in the nucleus. Figure 8.17 overleaf shows how, as the atomic number, Z, increases, the number of stable isotopes diverges from the line $Z = N$.

DID YOU KNOW?

The discovery of element 118 is an example of how important it is for scientific experiments to be repeatable by others. The discovery of element 118 was first reported by a team of scientists at Berkeley Lab in 2000, but in 2002 they retracted their paper after several confirmation experiments failed to reproduce the results.

The discovery announced in 2006 was a result of collaboration between scientists at the Lawrence Livermore National Laboratory in the USA and from Dubna, the Joint Institute for Nuclear Research in Russia. The discovery would only be confirmed after other groups had reproduced it.

© Not for Ethiopia Not to be republished

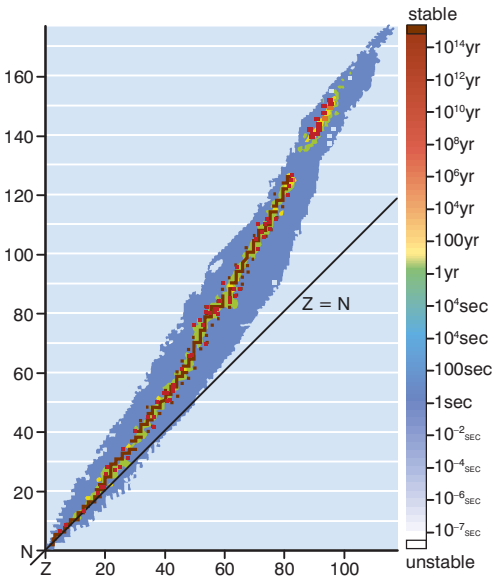


Figure 8.17

Nuclear stability is also determined by the **binding energy** per nucleon. Binding energy is defined as the energy required to disassemble a nucleus into the same number of free unbound protons and neutrons as it is composed of, in such as way that the particles are distant enough from each other so that the strong nuclear force can no longer cause the particles to interact. The net binding energy of a nucleus is that of nuclear attraction, minus the disruptive energy of the electrostatic force. Any system will always try and move to a state of lower energy (or more stable state).

As nuclei get heavier than helium, their net binding energy per nucleon (which can be found by calculating the difference in mass between the nucleus and the sum of the masses of the nucleons of which it is composed) grows more and more slowly and reaches its peak at iron, as shown in Figure 8.18.

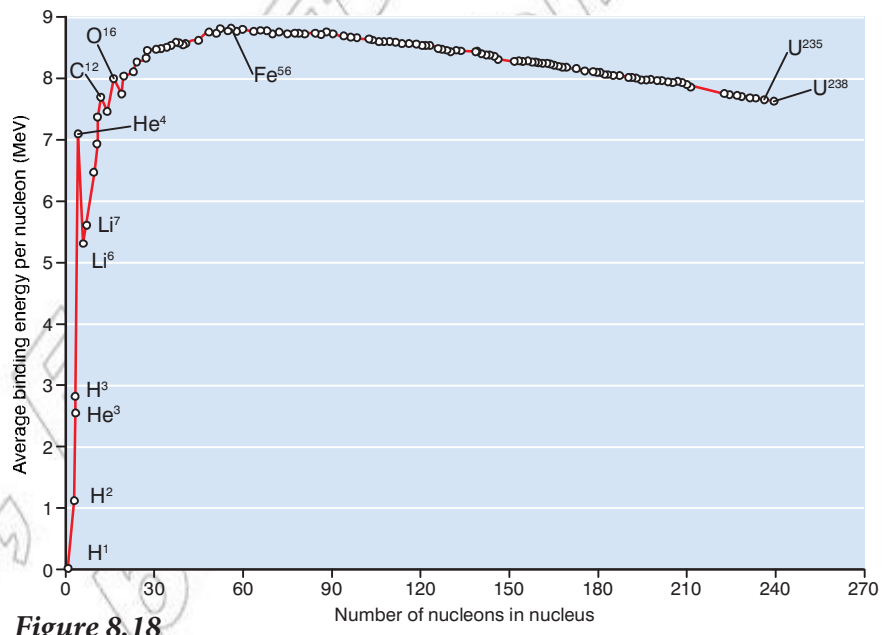


Figure 8.18

KEY WORDS

binding energy *the energy required to disassemble a nucleus into the same number of free unbound protons and neutrons as it is composed of, in such as way that the particles are distant enough from each other so that the strong nuclear force can no longer cause the particles to interact*

radioactivity *the process by which a nucleus will reach a lower energy state and thus become more stable by emitting particles*

As nucleons are added, the total binding energy increases but so does the total disruptive energy of the electrostatic forces and, once nuclei are heavier than that of iron, the increase in disruptive energy has more effect than the increase in binding energy. To reduce the disruptive energy, the weak interaction allows more neutrons to be added so that the number of neutrons exceeds the number of protons. However, at some stage the only way for the nucleus to reach a lower energy state (one that is more stable) will be to emit particles. This is the process we call **radioactivity**.

Radioactivity

If a nucleus emits particles and therefore loses energy in order to become more stable, then the nucleus has been split into two or more parts in the process known as **nuclear fission**. In other words, an atom of one type, the parent nuclide, transforms into an atom of another type, called the daughter nuclide, together with some form

© MoE Not to be republished

of radiation. According to quantum mechanics, it is impossible to predict precisely when a given atom will decay so radioactive emissions occur randomly over space. However, as we shall see on page 337, when a large number of similar atoms decay, the average decay rate can be predicted. The decay of nuclei is independent of conditions outside the nucleus – it is solely governed by the energy state of the nucleus.

Nuclear fusion is the process by which two nuclides are fused together to form a new nuclide. This process requires high temperature and pressure. It occurs naturally in stars but research into artificial nuclear fusion is ongoing.

KEY WORDS

nuclear fission *the process in which the nucleus becomes more stable by splitting into two or more parts and emitting particles*

nuclear fusion *the process by which two nuclides are fused together to form a new nuclide*

Types of nuclear radiation

There are three types of nuclear radiation: alpha (α), beta (β) and gamma (γ) radiation. Each of these comes about through a different process in the decaying nucleus, each one is composed of different particles and each one has different properties.

When a nuclear decay occurs, the particle emitted will leave the nucleus with a certain amount of kinetic energy. As the particle travels, it will ionize particles in its path, losing a small amount of kinetic energy at each ionization. When it has transferred all its kinetic energy, it will stop and will be absorbed by the substance it is in at that moment.

An alpha particle has two protons and two neutrons, the same as a helium nucleus. It can be written as ${}^{4}_{2}\alpha$. It is a relatively large particle with a significant positive charge of +2, so it ionises a lot. As it does so it loses its kinetic energy quickly and is easily absorbed. When it has travelled a few centimetres in air it is absorbed and it is completely blocked by paper or skin.

Beta particles are emitted from the nucleus when a neutron decays into a proton. It is a high speed particle, with a negative charge and little mass. It can be written as ${}^{0}_{-1}\beta$. Because it is much smaller than an alpha particle and has a single negative charge, it is much less ionizing than alpha particles and so it can penetrate much further. Several metres of air, or a thin sheet of aluminium, are needed to absorb beta particles.

Gamma rays are high energy, high frequency, electromagnetic radiation. They have no charge and no mass so they rarely interact with particles in their path, so they are the least ionizing of the three radiations. They are never completely absorbed, although their energy can be significantly reduced by several centimetres of lead, or several metres of concrete. If the energy is reduced to a safe level, gamma rays are often said to have been absorbed.

Radioactive sources need to be handled with extreme care. Safety precautions must be taken when using them (see box). Ionising radiations can interact with human cells. There may be so much ionization that cells die as a result. Where there is less ionisation,

© MOEP FET Physical Sciences Not to be republished

the molecules of DNA in the cell may change slightly, which could cause the cells to have an increased tendency to become cancerous. Because the radiations ionize to different extents, the hazard level is different for each one. The hazards are summarised in the table.

Type of radiation	Inside body	Outside body
alpha	Highly ionising – very dangerous radiation poisoning and cancer possible	Absorbed by surface layer of dead skin cells – no danger
beta	Moderate ionisation and danger should be minimised	Moderate so exposure ionisation and danger, close exposure should be minimised
gamma	minimal ionisation, cancer danger from long-term exposure	minimal ionisation – cancer danger from long–term exposure

Think about this...

Which of the three radiations do you think is most suitable for medical uses?

Safety precautions when using radioactive sources

Radioactive sources which are used in school are usually very weak.

They can only be used in the presence of an authorised teacher.

They are kept in a sealed container except when they are being used in an experiment or demonstration.

They are immediately returned to the container when the experiment or demonstration is finished.

When using the radioactive source it should be

1. Handled with tongs or forceps, never with bare hands.
2. Kept at arm's length, pointing away from the body.
3. Always kept as far as possible from the eyes.

Hands must be washed after the experiment and definitely before eating.

Activity 8.7: Predict the effect of magnetic and electric fields on the motion of alpha, beta and gamma rays

Work in a small group. Based on what you know about alpha, beta and gamma radiations, discuss what effect magnetic and electric fields will have on the motion of alpha, beta and gamma rays. Justify your reasoning.

Activity 8.8: Simulating nuclear reactions

Use marbles accelerated down a sloping aluminium channel into a saucer to simulate nuclear reactions produced by high-speed particles. The marbles in the saucer are analogous to target nuclei. Try using several marbles of different sizes. Show the effect of speed by launching the marbles from different heights and angles and the effect of mass and increased momentum of projectiles by using marbles of different sizes. Marbles coming down the slope with sufficient momentum can eject one of more marbles from the saucer. Discuss how this is analogous to bombarding nuclei with ever increasing mass: a proton, deuteron (two protons) and an alpha particle.

Activity 8.9: Penetrating power of alpha, beta and gamma radiation

Work in a small group. **Before you begin, make sure that you know the safety precautions that you must follow when using radioactive materials.**

Set up the equipment as shown in Figure 8.19.

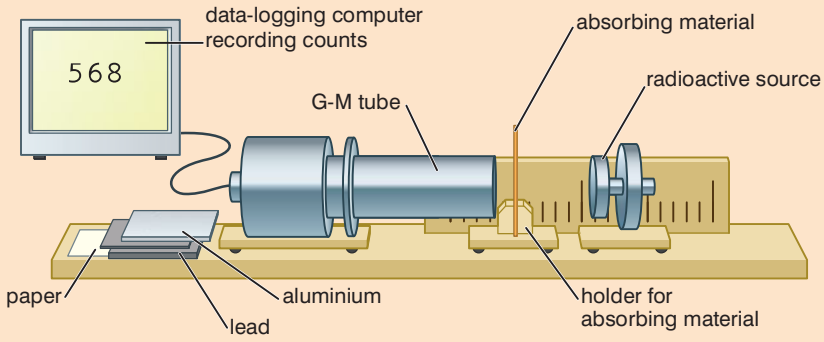


Figure 8.19

The source of radiation can emit alpha, beta or gamma rays. The Geiger–Muller tube will detect all three types of radiation. Place absorber sheets which progressively increase in density between the source and the detector, and record the average count rate in each case.

Note that, in order to remove all risk of exposure to radiation hazards, this experiment is often carried out using computer software.

Activity 8.10: Research the common detectors for α–particle, β–particle and γ–rays

In a small group, research the various forms of detectors for nuclear radiation. Prepare a summary of your research to present to the rest of your class. Choose an appropriate format for your presentation.

The relationship between mass and energy

One of Einstein's most important theories suggests that energy and mass are related by the equation

$E = mc^2$

where E is energy, m is mass and c is the speed of light (3×10^8 m/s).

On page 330, we learnt about the binding energy, that is, the energy required to hold the nucleus together. Any nuclear reaction which increases the binding energy per nucleon will give out energy. Using Einstein's equation, we can see that a change in energy will be associated with a change in mass. In fact, we find that there is a difference between the mass of a nucleus that we calculate by adding up the masses of its constituent protons and neutrons, and the measured mass of the nucleus. This difference is called the mass deficit, Δm.

© Not to

When calculating the binding energy, we often use atomic mass units for the masses of subatomic particles.

Particle	Mass (atomic mass units, u)	Mass (kg)
proton	1.007 276	$1.672\,623 \times 10^{-27}$
neutron	1.008 665	$1.674\,929 \times 10^{-27}$
electron	0.000 548 58	$9.109\,390 \times 10^{-31}$

If you use atomic mass units in nuclear energy calculations, then you multiply Δm (in atomic mass units) by 931.5 to give your result in MeV.

Worked example 8.8

a) Find the mass deficit for a carbon-12 nucleus.

b) Use this mass deficit to calculate the binding energy for a carbon-12 nucleus in joules.

c) Use this mass deficit to calculate the binding energy for a carbon-12 nucleus in electronvolts.

a) Use Figure 8.20 for this part of the question.



measured mass of C-12 atom = 12.000 000 u
mass of C-12 nucleus is atomic mass minus electron mass
= 12.000 000 – (6 × 0.000 548 58)
= 11.996 709 u



total mass of 6 neutrons
= 6 × 1.008 665
= 6.051 99 u



total mass of 6 protons
= 6 × 1.007 726
= 6.043 656 u

total mass of nucleons in C-12 nucleus
= 6.051 99 + 6.043 656
= 12.095 646 u

mass deficit for carbon-12:
Δm = 12.095 646 – 11.996 709 = 0.098 937 u
Δm = 0.098 937 × 1.66 × 10^{-27} = 1.642 354 × 10^{-28} kg

Figure 8.20

b)

ΔE (J)	Δm (kg)	c (m/s)
?	$1.642\,354 \times 10^{-28}$	3×10^{8}

Use $\Delta E = \Delta m c^2$

$= 1.642\,354 \times 10^{-28} \times (3 \times 108)^2$

$= 1.478 \times 10^{-11}$ J

c)

ΔE (MeV)	Δm (u)	c (m/s)
?	$1.642\,354 \times 10^{-28}$	3×10^{8}

Use $\Delta E = \Delta m \times 931.5$

$= 0.098\,937 \times 931.5$

$= 92.2$ MeV

© MoE Not to

Representing and interpreting nuclear reactions

We can represent nuclear reactions in the form of equations such as

$^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}\beta$

This equation represents the decay of a carbon-14 nucleus leaving a daughter nitrogen nucleus and emitting a beta particle in the process. Notice that the equation is 'balanced' in the sense that the total mass number and atomic number is the same on both sides of the arrow. This must always be the case. We know that energy is conserved in the reaction, so the energy released when a carbon-14 nucleus decays to a nitrogen nucleus is emitted in the form of a beta particle.

Worked example 8.9

Complete the following equation which represents a nuclear fission reaction and find the value of X.

$^{235}U + {}^{1}_{0}n \rightarrow {}^{X}Rb + {}^{139}Cs + 2\,{}^{1}_{0}n + \text{energy}$

The equation is balanced on both sides so the total mass number on the left side must be the same as the total mass number on the right side

$235 + 1 = X + 139 + 2$

$236 = X + 141$

$236 - 141 = X$

$= 95$

We can combine nuclear reaction equations and data about the mass of the elements on both sides of the equation to find out how much energy is released per fission.

Worked example 8.10

Calculate the energy released in the following fission reaction. Give your answer in MeV.

$^{235}U + {}^{1}_{0}n \rightarrow {}^{141}Ba + {}^{92}Kr + 3\,{}^{1}_{0}n + \text{energy}$

The data you need are:

mass of U-235 is 235.0439 u

mass of Ba-141 is 140.9144 u

mass of Kr-92 is 91.9262 u

mass of $^{1}_{0}n$ is 1.008 665 u.

Consider mass on each side of the equation

$235.0439\text{ u} + 1.008\,66\text{ u} \rightarrow 140.9144\text{ u} + 91.9262\text{ u} + (3 \times 1.008\,665\text{ u})$

$236.05256\text{ u} \rightarrow 235.866595\text{ u}$

$\Delta m = (236.05256 - 235.866595)\text{ u}$

$= 0.185\,965\text{ u}$

Energy released $= 0.185\,96 \times 931.5\text{ MeV}$

$= 173\text{ MeV}$

Ethiopia published

Radioactive half-life

We know that radioactive decay is a random process. For every second, there is a probability that a nucleus will decay which is called the decay constant and given the symbol λ. If we have a sample of the nuclei, the probability of decay will determine the fraction of the sample that will decay. Of course, if the sample is larger, then more nuclei will decay in each second. This means that the activity (A) (the number decaying per second) is proportional to the number of nuclei in the sample, N. Mathematically we write this as

$$A = -\lambda N$$
$$\frac{dN}{dt} = -\lambda N$$

There is a minus sign in the formula because the number of nuclei in the sample decreases with time but in practice we ignore the minus sign when we use the formula. The units for activity are bequerel (Bq).

Worked example 8.11

What is the activity of a sample of 100 million atoms of carbon-14? The decay constant, λ, is 3.84×10^{-12} s^{-1}.

$\frac{dN}{dt}$ (Bq)	λ (s^{-1})	N
?	3.84×10^{-12}	100×10^{6}

Use $\frac{dN}{dt} = -\lambda N$

$= (3.84 \times 10^{-12}) \times (100 \times 10^{6})$

$= 3.84 \times 10^{-4}$ Bq

The formula for the rate of decay of nuclei in a sample is a differential equation that can be solved to give a formula for the number of nuclei remaining in a sample, N, after a fixed time, t

$$N = N_0 e^{-\lambda t}$$

where N_0 is the initial number of nuclei within a sample and λ is the decay constant.

© MoE Ethiopia Not to be republished

Worked example 8.12

If the sample of 100 million carbon-14 atoms in worked example 8.11 were left for 250 years, how many carbon-14 atoms would remain?

N	N_0	λ (s^{-1})	t (s)
?	100×10^6	3.84×10^{-12}	$250 \times 365 \times 24 \times 60 \times 60 = 7.884 \times 10^9$

Use $N = N_0 e^{-\lambda t}$

$= 100 \times 10^6 \times e^{-7.884 \times 10^9 \times 3.84 \times 10^{-12}}$

$= 100 \times 10^6 \times e^{-0.03027456}$

$= 9.701 \times 10^7$ atoms

KEY WORDS

half-life *the time taken for half the atoms of a given nuclide within a sample to decay*

We know that the activity of a sample of radioactive nuclei decreases over time and that the activity depends on the number of nuclei present. The rate at which the activity decreases depends on the particular isotope that is decaying. A measure of this rate of decrease of activity is called the **half-life**, $t_{1/2}$. The half life can be defined as the time taken for half the atoms of a given nuclide within a sample to decay.

We can find a mathematical expression for the half-life by putting $N = \frac{1}{2}N_0$ into the decay equation

$$N = N_0 e^{-\lambda t}$$

$$\frac{1}{2} N_0 = N_0 e^{-\lambda t_{1/2}}$$

$$\frac{1}{2} = e^{-\lambda t_{1/2}}$$

$$\ln \frac{1}{2} = -\lambda t_{1/2}$$

$$-\ln 2 = -\lambda t_{1/2}$$

$$t_{1/2} = \frac{\ln 2}{\lambda}$$

$$\lambda = \frac{\ln 2}{t_{1/2}}$$

© MoE, FDRE Ethiopia Not to be republished

Worked example 8.13

What is the half-life of carbon-14?

λ (s^{-1})	ln 2	$t_{1/2}$ (s)
3.84×10^{-12}	0.6931	?

Use $t_{1/2} = \dfrac{0.6931}{3.84 \times 10^{-12}}$

$= 1.81 \times 10^{11}$ s

$= 5.027777778 \times 10^{7}$ hours

$= 2.094907407 \times 10^{6}$ days

$= 5739$ years

If you were to carry out an experiment to measure the half-life of a radioactive substance, you would measure its activity over time. Activity is proportional to the number of nuclei present and so, when the activity is plotted against time, the shape of the curve is exponential decay as shown in Figure 8.21.

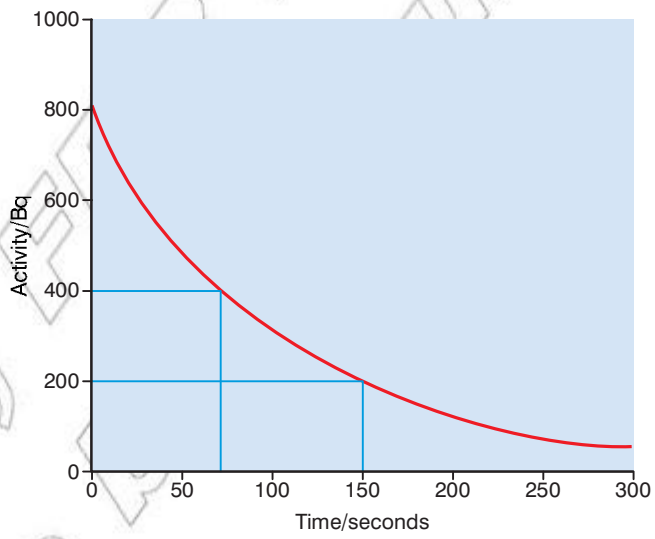


Figure 8.21

The activity, A, follows the equation

$A = A_0 e^{-\lambda t}$

The graph can be used to find the half-life of the substance by finding the time it takes for the activity to halve. On Figure 8.21, you can see that the time taken for the activity to drop from 800 Bq to 400 Bq was 70 s and the time taken to drop from 400 Bq to 200 Bq was 80 s. This gives an average for the half-life of 75 s. The time interval is not identical each time because of the random nature of radioactive decay and experimental and graphing errors. For this reason, several values need to be taken for the half-life and then these should be averaged, as above.

© MoE, Not to be republished

Worked example 8.14

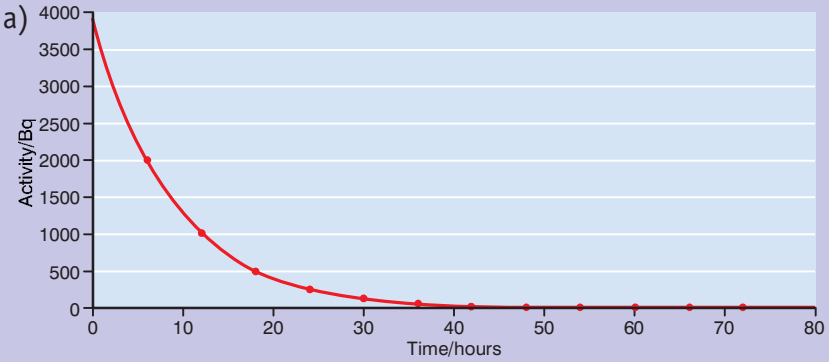
Here is a table showing the activity of a sample of technetium-99, a gamma emitter which is often used in medical investigations.

Activity (Bq)	Time (hours)
4000	0
2000	6
1000	12
500	18
250	24
125	30
75	36

a) Plot a graph to show this data.

b) Work out the half life for technetium-99.

a)



b) The graph shows that the activity falls from 4000 Bq to 2000 Bq in 6 hours, from 2000 Bq to 1000 Bq in 6 hours and from 1000 Bq to 500 Bq in 6 hours. The half-life for this substance is therefore 6 hours.

Uses of radioactive isotopes

Radioactive substances have many applications. Radiation is used to treat cancer – the radiation destroys the cancerous cells, while treatment is designed to protect non-cancerous cells. An example is the use of iodine-131, in the form of iodine chloride, which is used to treat thyroid cancer. The patient takes the radioactive substance orally and then the chemical travels to the thyroid and the radiation treats the disease.

The most commonly used radioactive isotope for medical applications is technetium-99. This is the daughter nuclide from the decay of molybdenum-99, which is itself produced from the fission on uranium-235 in nuclear reactors. Technetium-99 emits 140 keV gamma radiation so it is relatively low energy for gamma radiation and is therefore less likely to produce damaging ionisation in the body but is energetic enough to be detected outside the body. Technetium-99 is used for gamma ray scanning to produce images

Activity 8.11: Uses of radioisotopes in dating archaeological samples

In a small group, research the uses of radioisotopes in dating archaeological samples.

of the body, and also as a tracer to check the function of different organs of the body, including the bone marrow, brain and heart.

The radioactive isotope plutonium-238, which emits alpha radiation, is commonly used in atomic power supplies, such as those required for space travel.

Radioactive isotopes are also used in nuclear power stations to generate electricity. An objection raised against nuclear power is the problem of disposing of the waste material, which we shall discuss next.

Activity 8.12: Research nuclear power

Work in a small group to research one of the following topics.

- The fraction of energy generated from nuclear power in Africa and the rest of the world
- Peaceful uses of nuclear radiation in Ethiopia and Africa
- Nuclear facilities in Africa

Present your findings to the rest of your class in a form of your choice.

The problems posed by nuclear waste

In your research for Activity 8.12, you may have come across some of the problems posed by nuclear waste. One of the main difficulties is that the isotopes used in nuclear power stations typically have very long half lives. Plutonium-239 has a half-life of 24 100 years; in contrast plutonium-238 has a half-life of 88 years.

Spent nuclear fuel is the most important source of waste from nuclear power stations and is mainly unconverted uranium. About 3% of it is fission products from nuclear reactions. The actinides (uranium, plutonium, and curium) are responsible for the bulk of the long-term radioactivity, whereas the fission products are responsible for the bulk of the short-term radioactivity.

After about 5 percent of a nuclear fuel rod has reacted inside a nuclear reactor, that rod is no longer able to be used as fuel (due to the build-up of fission products). Scientists are experimenting on methods for reusing these rods in order to reduce waste and use the remaining actinides as fuel (large-scale reprocessing is being used in a number of countries).

A typical 1000- MW nuclear reactor produces approximately 20 m^3 (about 27 tonnes) of spent nuclear fuel each year (but this reduces to 3 m^3 if the waste is reprocessed). The remaining waste will be substantially radioactive for at least 300 years.

Spent nuclear fuel is initially very highly radioactive and so must be handled with great care. It becomes significantly less radioactive over the course of thousands of years of time. Some scientists believe that, after 10 000 years of radioactive decay, the spent nuclear fuel will no longer pose a threat to public health and safety.

© MoE Not to be republished

When they are first extracted from the reactor, spent fuel rods are stored in shielded basins of water (spent fuel pools), usually located on-site. The water provides both cooling for the still-decaying fission products, and shielding from the continuing radioactivity. After about five years, the now cooler, less radioactive fuel is typically moved to a dry-storage facility or dry cask storage, where the fuel is stored in steel and concrete containers.

An article published in 2007 states 'Today we stock containers of waste because currently scientists don't know how to reduce or eliminate the toxicity, but maybe in 100 years perhaps scientists will... Nuclear waste is an enormously difficult political problem which to date no country has solved. It is, in a sense, the Achilles heel of the nuclear industry.'

Despite all this, you should be aware that in countries with nuclear power, radioactive wastes comprise less than 1% of total industrial toxic wastes, much of which remains hazardous indefinitely. Overall, nuclear power produces far less waste material by volume than fossil-fuel based power plants. Coal-burning plants are particularly noted for producing large amounts of toxic and mildly radioactive ash due to concentrating naturally occurring metals and mildly radioactive material from the coal. It has been stated in a recent report from Oak Ridge National Laboratory that coal power actually results in more radioactivity being released into the environment than nuclear power operation, and that the population effective dose equivalent from radiation from coal plants is 100 times as much as from ideal operation of nuclear plants. Another factor is that although coal ash is much less radioactive than nuclear waste, ash is released directly into the environment, whereas nuclear plants use shielding to protect the environment from the irradiated reactor vessel, fuel rods, and any radioactive waste on site.

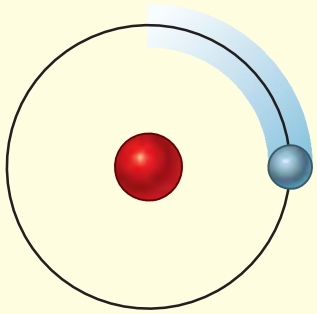
There is no easy answer to the issues and the debate will certainly continue for many years.

© FDRE Ethiopia MoE Not to be republished

Summary

In this section you have learnt that:

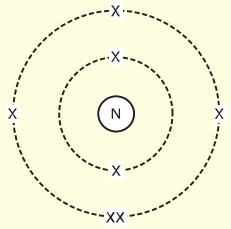
- Rutherford's model of atom is as shown here.



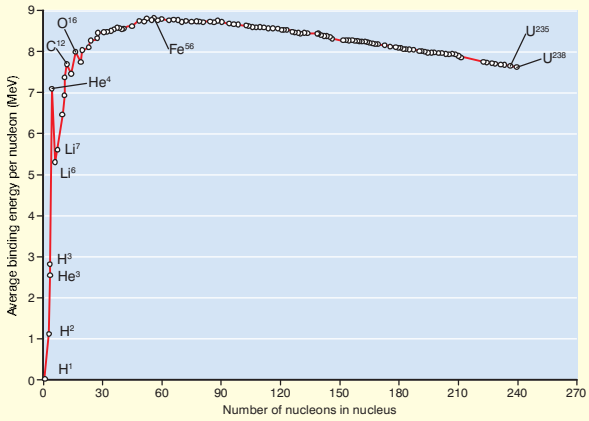
Rutherford's nuclear model of the atom: All the positive charge and most of the mass is concentrated in a tiny central nucleus. Most of the atom is empty space, and electrons orbit at the edge.

- Niels Bohr proposed that specific energy levels existed within an atom's structure and that electrons move in circular orbits around the nucleus. In his model, electrons that are closer to the nucleus have a lower energy state than those that are further away.

- Bohr's model means that electrons can only exist at specific energy states, and will not be found with energies between those levels.
- The change in energy of an atom can be calculated using the relation $\Delta E = E_f - E_i$
- The structure of simple atoms can be represented using simple diagrams like this.



- Isotopes of an element have the same atomic number (Z) but a different mass number (A). The relationship $A = Z + N$ means that isotopes of an element must therefore have different numbers of neutrons in their nuclei, but the same number of protons.
- The charge and mass of the electron is approximately $\frac{1}{1836}$ the charge and mass of the proton.
- The nuclear force is a very strong force that holds the particles in a nucleus together.
- The strong force is the same between any two nucleons and acts over a very short range.
- The radius of a nucleus and its mass number are related mathematically by the relationship $R = (1.2 \times 10^{-15}\ \text{m})A^{1/3}$.
- Experimental results show that the radii of atoms vary from 35×10^{-12} m for hydrogen atoms where A is 1, to 175×10^{-12} m for americium atoms where A is 95.
- When nuclei of atoms are ordered according to atomic number and number of nucleons, the following properties are observed.
  - For the lighter nuclei, if we look at the most common isotope, N is approximately equal to Z.
  - As we get to heavier nuclei, past Z = 20, we begin to see N considerably greater than Z. As nuclei get heavier this becomes more apparent.
  - Bismuth is the heaviest stable nucleus. Heavier nuclei exist but they are all unstable – they undergo certain spontaneous changes which we observe as radioactivity. Nuclei from Z = 84 (polonium) to 92 (uranium) are found in nature (on Earth) and all their isotopes are radioactive.
  - Nuclei heavier than uranium exist but they are all artificial – they have been created by scientists in laboratories. The heaviest known nucleus has Z = 118. It was produced in 2006.
- **Binding energy** is the energy required to disassemble a nucleus into the same number of free unbound protons and neutrons as it is composed of, in such as way that the particles are distant enough from each other so that the strong nuclear force can no longer cause the particles to interact
- As nuclei get heavier than helium, their net binding energy per nucleon (which can be found by calculating the difference in mass between the nucleus and the sum of the masses of the nucleons of which it is composed) grows more and more slowly and reaches its peak at iron, as shown in the diagram.



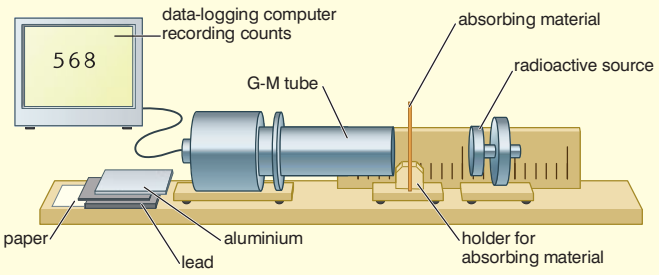
- As nucleons are added, the total binding energy increases but so does the total disruptive energy of the electrostatic forces and, once nuclei are heavier than iron, the increase in disruptive energy has more effect than the increase in binding energy. To reduce the disruptive energy, the weak interaction allows more neutrons to be added so that the number of neutrons

exceeds the number of protons. However, at some stage the only way for the nucleus to reach a lower energy state (one that is more stable) will be to emit particles by **radioactivity**.

- **Nuclear fission** is the nucleus becoming more stable by splitting into two or more parts and emitting particles. An atom of one type, the parent nuclide, transforms into an atom of another type, called the daughter nuclide, together with some form of radiation. According to quantum mechanics, it is impossible to predict precisely when a given atom will decay so radioactive emissions occur randomly over space. The nuclear decay process is independent of conditions outside the nucleus.
- **Nuclear fusion** the process by which two nuclides are fused together to form a new nuclide. This process requires high temperature and pressure. It occurs naturally in stars but research into artificial nuclear fusion is ongoing.
- The three types of emissions from radioactive substances are alpha, beta and gamma rays.
- Because of the ionizing effects of the radiation, safety measures are needed when handling and using radio-isotopes.
- Radioactive sources which are used in school are usually very weak.
- They can only be used in the presence of an authorised teacher.
- They are kept in a sealed container except when they are being used in an experiment or demonstration.
- They are immediately returned to the container when the experiment or demonstration is finished.
- When using the radioactive source it should be
  1. Handled with tongs or forceps, never with bare hands.
  2. Kept at arm's length, pointing away from the body.
  3. Always kept as far as possible from the eyes.
- Hands must be washed after the experiment and definitely before eating.

Radiation	Nature	Relative ionizing effect	Relative penetrating power	Effect of electric field on radiation	Effect of magnetic field on radiation	Common detector
alpha	$^{4}_{2}He$ nucleus	Highly ionising	A few centimetres in air, completely blocked by paper and skin	Deflects	Deflects	Geiger–Muller tube
Beta	Negative charge, mass of electron	Less ionising than alpha	Several metres of air, thin sheet of aluminium will absorb	Deflects in opposite direction to alpha	Deflects in opposite direction to alpha	Geiger–Muller tube
Gamma	No mass, electromagnetic radiation	Least ionising radiation	Never completely absorbed but energy can be significantly reduced by several centimetres of lead or several metres of concrete	No charge so no deflection	No charge so no deflection	Geiger–Muller tube

- The range of alpha, beta and gamma radiation in various media can be compared using the equipment shown here.



- The release of energy in a nuclear reaction is associated with a change in mass according to Einstein's equation $E = mc^2$ where E is the energy released, m is the change in mass and c is the speed of light. This equation may also be used with atomic mass units where it reduces to $E = (931.5 \times \Delta m)$ MeV
- Nuclear reactions can be represented in the form ${}^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}\beta$
- This equation represents the decay of a carbon–14 nucleus leaving a daughter nitrogen nucleus and emitting a beta particle in the process. The equation is 'balanced' in the sense that the total mass number and atomic number is the same on both sides of the arrow.
- **Half-life** is the time taken for half the atoms of a given nuclide within a sample to decay

$$t_{1/2} = \frac{\ln 2}{\lambda}$$

$$\lambda = \frac{\ln 2}{t_{1/2}}$$

- Graphs of random decay show that such processes have a constant half-life.
- Uses of radioactive isotopes include: medical treatment and diagnosis, to date archaeological samples and to generate electricity.
- The problems posed by nuclear waste are a result of the very long half lives of the isotopes used to generate electricity, and the safe storage of the waste.

Review questions

1. Describe a) Rutherford's model of the atom, b) Bohr's model of the atom.
2. Suppose that an atom has two energy levels of value E_1 and E_2 above the ground-state level, taken as zero energy. Photons are emitted from the atom at wavelengths 620 nm (red), 540 nm (green) and 290 nm (uv).
   a) Sketch the energy level diagram.
   b) Add and label the three transitions causing the emitted light.
   c) Calculate the energy levels which will give these three spectral lines in joules.
3. Represent diagrammatically the structure of a carbon-6 atom.
4. What is meant by the term isotope?
5. How does the charge and mass of the electron compare with the charge and mass of the proton?

© MoE; Not to be r

6. a) What is the strong nuclear force?
   b) What are some important properties of the strong force?
7. a) How are the radius and the mass number related mathematically?
   b) What is the range for sizes of atomic nuclei?
8. State nuclear properties.
9. How is nuclear stability determined by binding energy per nucleon?
10. a) How is radioactivity associated with nuclear instability?
    b) What is the difference between nuclear fission and nuclear fusion?
11. a) Identify the nature of the three types of emissions from radioactive substances.
    b) Draw a table to distinguish between the three kinds of emissions in terms of their nature, relative ionizing effect, relative penetrating power, the effect of magnetic and electric fields on the radiation, and the common detectors.
    c) Describe the need for safety measures in handling and using radioisotopes.
    d) Describe experiments to compare the range of alpha, beta and gamma radiation in various media.
12. How is the release of energy in a nuclear reaction associated with a change in mass?
13. Caesium-137 is a by-product of nuclear fission within a nuclear reactor.
    a) Copy and complete this equation which describes the production of $^{137}_{55}Cs$.

    $$^{235}_{92}U + {}^{1}_{0}n \rightarrow {}^{137}_{55}Cs + {}^{95}Rb + {}^{1}_{0}n$$

    b) The half-life of is 30 years. When the fuel rods are removed from a nuclear reactor core, the total activity of caesium-137 is 5.8×10^{15} Bq. After how many years will this have fallen to 1.6×10^{6} Bq?
    c) Comment on the problems of storage of the fuel rods over this time period.
14. State some uses of radioactive isotopes.

© MoE FDRE Ethiopia Not to be republished

End of unit questions

1. Explain the meaning of the term 'black body'.
2. Radiation of wavelength 290 nm falls on a sample of magnesium. Photoelectrons are emitted. Calculate the energy of these electrons. Take the Planck constant to be 6.63×10^{-34} J s and the speed of light to be 3×10^{8} m/s.
3. An ultraviolet lamp is used to illuminate a clean sodium surface and photoelectrons are emitted. The stopping potential of the photoelectrons is –1.92 V. The work function energy of sodium is 2.36 eV. Calculate the wavelength of the UV radiation. Take the value of the Planck constant to be 6.63×10^{-34} J s.
4. Electrons are accelerated through a potential difference of 7500 V before striking a layer of graphite and being diffracted. The mass of an electron is 9.11×10^{-31} kg. Planck's constant is 6.63×10^{-34} J s.

   Calculate

   a) the speed of the electrons when they hit the layer of graphite

   b) the momentum of the electrons

   c) the wavelength of the electrons.
5. State Heisenberg's uncertainty principle.
6. Light of wavelength 2.5×10^{-11} m is used to measure the position of an electron of wavelength 1.54×10^{-11} m. Find the uncertainty in the position of the electron. Planck's constant is 6.63×10^{-34} J s.
7. How does Bohr's model of the atom refine Rutherford's model?
8. a) There is a dark line in the Sun's spectrum at 588 nm. Calculate the energy of a photon with this wavelength in joules. Planck's constant is 6.63×10^{-34} J s.

   b) A helium atom as energy levels at -1.59×10^{-19} J, -2.42×10^{-19} J, -3.00×10^{-19} J, -5.80×10^{-19} J, -7.64×10^{-19} J. Explain, with reference to these energy levels, how the dark line in the Sun's spectrum at 588 nm may be due to the presence of helium in the gases which surround the Sun.
9. a) Represent diagrammatically the structure of a chlorine-35 atom. The atomic number of chlorine is 17.

   b) Use the example of chlorine to explain the term isotope.
10. List some important properties of the nuclear strong force.
11. State nuclear properties.
12. Explain how nuclear instability and binding energy can lead to the release of energy.

© MoE. For Ethiopia Not to be republished

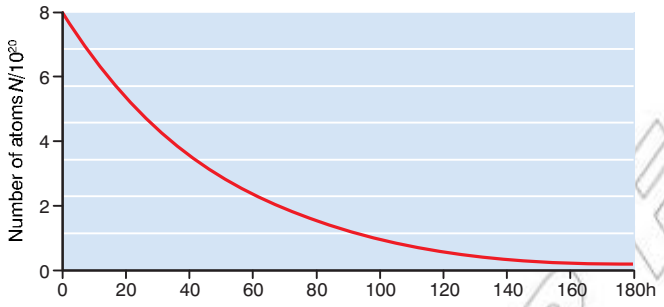
13. Compare alpha and beta emissions in terms of their nature, relative ionising effect, relative penetrating power, the effect of magnetic and electric fields on the radiation, and the common detectors.

14. Radon-220 is a radioactive gas which decays by alpha emission to polonium-216. The atomic number for polonium is 84.

    a) Write a nuclear reaction equation to describe this decay.

    b) The half-life of this decay is about 1 minute. Describe an experiment that you could perform to check this half-life value.

    c) The graph shows the decay of a radioactive nuclide.



    Determine the half-life of this nuclide.

    d) Use your value of half-life to calculate the decay constant λ of this radionuclide.

15. Explain carbon dating.

© MoE, FDR Ethiopia
Not to be republished

Index

a.c. circuits 287–302, 304–9
  see also electric current
  a.c. current vs d.c. current 281–2
  capacitors 190
  power 304–9
  power factor 294
alternating current 287–302
  see also electric current
  capacitive circuits 290–2
  inductive circuits 292–4
  resistive circuits 289–90
  RLC circuits 300–302, 306–309
  root mean square (r.m.s.) values 287–8
a.c. generators 282–3
adiabatic process, bicycle pump 11–12
Ampere's law
  magnetism/magnetic fields 256–7
  solenoids 257–8
  toroids 259
angular frequency, simple harmonic motion (SHM) 61–3
astronomy, Doppler effect 104
atomic physics 311–44
  atom models 323–7
  atoms and nuclei 322–44
  black bodies 313
  Bohr model of the atom 324–5
  electronvolts 315
  Heisenberg's uncertainty principle 319–20
  mass/energy relationship 333–4
  matter/radiation 312–21
  nuclear properties 329
  nuclear radiation 331–3
  nuclear reactions 335
  nuclear stability 329–30
  nuclear waste 340–1
  photoelectric effect 313–16
  radioactive half-life 336–9
  radioactive isotopes 339–40
  radioactivity 330–1
  Rutherford's model of the atom 323–4
  strong nuclear force 327–8
  wave–particle duality of matter 317–19

beats, standing waves 94–5
bicycle pump, adiabatic process 11–12
Biot–Savart law 253
black bodies, atomic physics 313
Bohr model of the atom 324–5
Brownian motion 24
  diffusion 27–8

capacitance 174–6
capacitive circuits, a.c. 290–2
capacitors
  a.c. circuits 190, 290–2
  charging 188–90
  constructing 176–8
  dipoles 177
  discharging 184–8
  electrical energy density 183
  electrostatics 173–90
  Gauss's law 176
  Leyden jar 178–9
  in parallel 179
  in series 180
  uses 190–1
charged particle movement
  electric fields 150, 240–2
  magnetic fields 240–6
coherence, interference 131–3
conduction electrons, electric current 200–1
conductivity, electric current 201
constructive interference, wave motion 85–6
  corpuscular theory, reflection/refraction 117–19
coulomb, electric current 199–200
Coulomb's law, electrostatics 147–51
current density, electric current 203–5

Dalton's law of partial pressures 23
damping, oscillation 73–5
d.c. current, vs a.c. current 281–2
decibel, hearing 100–1
destructive interference, wave motion 86
dielectrics, electrostatics 176–8
diesel engines 42–4
diffraction
  diffraction grating 136–39
  single slit diffraction 133–6
  wave optics 120–4, 133–9
  Brownian motion 27–8
  Graham's law 28–9
dipoles
  capacitors 177
  electric fields 156–7
Doppler effect, wave motion 102–4
drift velocity, electric current 203–5

Earth, magnetism/magnetic fields 238–9, 260–3
eddy currents, electromagnetic induction 273–4
efficiency
  heat engines 34, 44–5
  second law of thermodynamics 31–6
electric current 196–230
  see also a.c. circuits; alternating current
  Biot–Savart law 253
  conduction electrons 200–1
  conductivity 201
  coulomb 199–200
  current density 203–5
  drift velocity 203–5
  electromotive force (e.m.f.) 205–7
  galvanometers 220–26
  Kirchoff's rules 214–19
  magnetic fields 247–55
  magnetic force, current-carrying conductors 247–55
  measuring instruments 220–6
  potential difference (p.d.) 205–8
  potentiometer 226–8
  resistance 201–2, 206–11
  resistivity 201–2
  voltage 205–8
  Wheatstone bridge 226–8
electric fields 143–60
  charged particles movement 155–6, 240–2
  dipoles 156–7
  electric potential 167–9
  vs gravitational fields 171
  parallel plates 154–5
  point charge 153–4
  strength 145
electric generators 282–6
electric potential
  electric fields 167–9
  electrical potential energy 169–70
  equipotentials 165–7
  point charge 164–5
electrical energy density, capacitors 183
electrical energy transmission 284–5
electromagnetic induction 268–80
  eddy currents 273–4
  induced e.m.f. 269–73, 275–6

© MOFED Ethiopia Not to be republished

laws 271–3
magnetic energy density 278
magnetic flux 268–9
mutual inductance 273, 274–5
self-inductance 274–5
solenoids 276–7
electromotive force (e.m.f.)
electric current 205–7
induced e.m.f. 269–73
electronvolts, atomic physics 315
electrostatics 141–92
capacitors 173–90
Coulomb's law 1497–51
electric charge 143–60
electric potential 162–70
electrostatic forces vs gravitational forces 152
force calculations 149–50
Gauss's law 152–3
Millikan's oil drop experiment 157–9
Stokes' law 158–9
energy/mass relationship, atomic physics 333–4
entropy, second law of thermodynamics 31–3
equipotentials, electric potential 165–7

first law of thermodynamics 9–20
forms 15–16
gases 16–20
internal energy 10–13
frequency, simple harmonic motion (SHM) 60

galvanometers
electric current 220–26
tangent galvanometers 261–2

gases
Brownian motion 24
Dalton's law of partial pressures 23
first law of thermodynamics 16–20
kinetic theory 21–30
laws 21–3
Gauss's law
capacitance 176
electrostatics 152–3
Graham's law of diffusion 28–29
gravitational fields vs electric fields 171
gravitational forces vs electrostatic forces 152

half-life, radioactive, atomic physics 336–9
harmonics, standing waves 92–4
hearing 97–104
decibel 100–1
heat engines 37–47
diesel engines 42–4
efficiency 34, 44–5
petrol engines 41–2
refrigerators 46–7
second law of thermodynamics 34
heat pumps 46–7
Heisenberg's uncertainty principle, atomic physics 319–20
Huygens's principle
diffraction 120–1
reflection/refraction 116–17
wave fronts 111–12

induced e.m.f.
electromagnetic induction 269–73, 275–6
inductance 275–6
inductive circuits, alternating current 292–4
inductors, alternating current 292–7, 298–302
intensity, sound 98–9
interference
coherence 131–3
constructive/destructive 85–6
interferometer 124–5
thin-film 125–7
wave motion 85–6
wave optics 120–33
Young's double slit experiment 128–32
internal energy, thermodynamics 5, 10–14

kinetic theory of gases, thermodynamics 21–30
Kirchoff's rules, electric current 214–19

lasers, Young's double slit experiment 132
Leyden jar, capacitors 178–9
longitudinal waves 81–2
loudness 98–101

magnetic energy density, electromagnetic induction 278
magnetic flux, electromagnetic induction 268–9
magnetism/magnetic fields 235–63
Ampere's law 256–7
Biot–Savart law 254
charged particles movement 240–6
circular motion of particles 244–5
current-carrying conductors 247–55
Earth 238–9, 260–3
electric current 247–55
Thompson's experiment 242–4
mass/energy relationship, atomic physics 333–4
mass–spring systems
periodic motion 55–6
time periods 67–9
matter/radiation, atomic physics 312–21
medical imaging, Doppler effect 104
Millikan's oil drop experiment, electrostatics 157–9
molar mass 4
moles 3–4
musical instruments, standing waves 87–94
mutual inductance, electromagnetic induction 273, 274–5

nuclear properties, atomic physics 329
nuclear radiation, atomic physics 331–3
nuclear reactions, atomic physics 335
nuclear stability, atomic physics 329–30
nuclear waste, atomic physics 340–1

organ pipes, standing waves 92–4
oscillation 53–6
see also periodic motion
damping 73–5
resonance 71–3

pendulums
periodic motion 54
time periods 69–70
periodic motion 53–78
see also wave motion
mass-spring systems 55–6
oscillation 53–6
pendulums 54
simple harmonic motion (SHM) 56–71, 75–8
petrol engines 41–2
phases of matter 7–8
photoelectric effect, atomic physics 313–16
potential difference (p.d.), electric current 205–8
potentiometers, electric current 228–30
power factor, a.c. circuits 294
power a.c. circuits 304–9

radar, Doppler effect 104
radiation/matter
atomic physics 312–21
nuclear radiation 331–3
radioactive half-life, atomic physics 336–9
radioactive isotopes, atomic physics 339–40

radioactivity, atomic physics 330–1
ray diagrams 110, 113
reflection
corpuscular theory 117–19
Huygens's principle 116
wave fronts 113–14
wave motion 87

© MoFET Editions
Not to be republished

refraction
corpuscular theory 117–19
Huygens's principle 117
wave fronts 114–15
refrigerators, heat engines 46–7
resistance, electric current 201–2, 206–11
resistive circuits, a.c. 289–90
resistivity, electric current 201–2
resistors, a.c. 289–90, 94–8, 300–2
resonance, oscillation 71–3
reversible/irreversible processes, second law of thermodynamics 35–6
RLC circuits, a.c. 300–2, 306–9
root mean square (r.m.s.) values, a.c. 287–8
Rutherford's model of the atom 323–4

second law of thermodynamics 1–6
self-inductance, electromagnetic induction 274–5
simple harmonic motion (SHM)
angular frequency 61–3
displacement 62–6
energy 76–9
frequency 60
periodic motion 56–71, 75–8
single slit diffraction 133–6
solenoids
Ampere's law 257–8
electromagnetic induction 276–7
sound 97–105
Doppler effect 102–4
intensity 98–9
speed of sound 101
standing waves 87–94
beats 94–5
harmonics 91–4
organ pipes 92–4
strings 87–92
wavelength 90–2
Stokes' law, electrostatics 158–9
strong nuclear force, atomic physics 327–8
superposition, wave motion 85

tangent galvanometers 261–2
temperature 6–7
thermal equilibrium 5–6
thermodynamics 1–47
entropy 31–3
first law of 9–20
kinetic theory of gases 21–30
second law of 31–6
temperature 6–7
thermal equilibrium 5–6
Thompson's experiment, magnetism 242–4
time periods 60
mass–spring systems 67–9
pendulums 69–70
toroids, Ampere's law 259
transformers 281–4
transverse waves 81
travelling waves 80–7

voltage, electric current 205–8

wave fronts 109–17
Huygens's principle 111–12, 116–17
reflection 113–16
refraction 114–17
wave motion 80–105
see also periodic motion
constructive interference 85–6
destructive interference 86
Doppler effect 102–4
interference 85–6
longitudinal waves 81–2
mathematical description 83–5
reflections of waves 87
speed 82
standing waves 87–94
superposition 85
transverse waves 81
travelling waves 80–7
wave optics 108–39
diffraction 120–4, 133–9
interference 120–33
reflection/refraction 113–17
wave fronts 109–17
wave-particle duality of matter, atomic physics 317–19
wavelength
standing waves 90–2
travelling waves 81–3
wave optics 109–10
Wheatstone bridge, electric current 226–8

Young's double slit experiment, interference 128–32

zeroth law, thermodynamics 5–6

© MoE-EDR Ethiopia
Not to be republished

© MoE, FDR Ethiopia
Not to be republished

© MoE, FDR Ethiopia
Not to be republished