



FHSST Authors

**The Free High School Science Texts:
Textbooks for High School Students
Studying the Sciences
Physical Science
Grade 12**

**Version 0.5
September 9, 2010**

Copyright 2007 "Free High School Science Texts"

Permission **is** granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".



STOP!!!!

Did you notice the **FREEDOMS** we've granted you?

Our copyright license is **different!** It grants freedoms rather than just imposing restrictions like all those other textbooks you probably own or use.

- We know people copy textbooks illegally but we would LOVE it if you copied our's - go ahead copy to your hearts content, **legally!**
- Publishers' revenue is generated by controlling the market, we don't want any money, go ahead, distribute our books far and wide - we DARE you!
- Ever wanted to change your textbook? Of course you have! Go ahead, change ours, make your own version, get your friends together, rip it apart and put it back together the way you like it. That's what we really want!
- Copy, modify, adapt, enhance, share, critique, adore, and contextualise. Do it all, do it with your colleagues, your friends, or alone but get involved! Together we can overcome the challenges our complex and diverse country presents.
- So what is the catch? The only thing you can't do is take this book, make a few changes and then tell others that they can't do the same with your changes. It's share and share-alike and we know you'll agree that is only fair.
- These books were written by volunteers who want to help support education, who want the facts to be freely available for teachers to copy, adapt and re-use. Thousands of hours went into making them and they are a gift to everyone in the education community.

FHSST Core Team

Mark Horner ; Samuel Halliday ; Sarah Blyth ; Rory Adams ; Spencer Wheaton

FHSST Editors

Jaynie Padayachee ; Joanne Boule ; Diana Mulcahy ; Annette Nell ; René Toerien ; Donovan Whitfield

FHSST Contributors

Sarah Abel ; Dr. Rory Adams ; Andrea Africa ; Ben Anhalt ; Prashant Arora ; Raymond Barbour ; Richard Baxter ; Tara Beckerling ; Tim van Beek ; Jennifer de Beyer ; Dr. Sarah Blyth ; Sebastian Bodenstein ; Martin Bongers ; Stephan Brandt ; Craig Brown ; Graeme Broster ; Deanne de Bude ; Richard Case ; Fanny Cherblanc ; Dr. Christine Chung ; Brett Cocks ; Andrew Craig ; Tim Crombie ; Dan Crytser ; Dr. Anne Dabrowski ; Laura Daniels ; Sean Dobbs ; Esmi Dreyer ; Matthew Duddy ; Fernando Durrell ; Dr. Dan Dwyer ; Frans van Eeden ; Alex Ellis ; Tom Ellis ; Giovanni Franzoni ; Ingrid von Glehn ; Tamara von Glehn ; Lindsay Glesener ; Kevin Godby ; Dr. Vanessa Godfrey ; Dr. Johan Gonzalez ; Hemant Gopal ; Dr. Stephanie Gould ; Umeshree Govender ; Heather Gray ; Lynn Greeff ; Dr. Tom Gutierrez ; Brooke Haag ; Kate Hadley ; Dr. Sam Halliday ; Asheena Hanuman ; Dr. Melanie Dymond Harper ; Dr. Nicholas Harrison ; Neil Hart ; Nicholas Hatcher ; Dr. William P. Heal ; Pierre van Heerden ; Dr. Fritha Hennessy ; Millie Hilgart ; Chris Holdsworth ; Dr. Benne Holwerda ; Dr. Mark Horner ; Mfandaizda Hove ; Robert Hovden ; Jennifer Hsieh ; Clare Johnson ; Luke Jordan ; Tana Joseph ; Dr. Fabian Jutz ; Dr. Lutz Kampmann ; Paul Kim ; Dr. Jennifer Klay ; Lara Kruger ; Sihle Kubheka ; Andrew Kubik ; Dr. Jannie Leach ; Dr. Marco van Leeuwen ; Dr. Tom Leinster ; Dr. Anton Machacek ; Dr. Komal Maheshwari ; Kosma von Maltitz ; Bryony Martin ; Nicole Masureik ; John Mathew ; Dr. Will Matthews ; JoEllen McBride ; Nikolai Meures ; Riana Meyer ; Filippo Miatto ; Jenny Miller ; Abdul Mirza ; Mapholo Modise ; Carla Moerdyk ; Asogan Moodaly ; Jothi Moodley ; David Myburgh ; Kamie Naidu ; Nolene Naidu ; Bridget Nash ; Tyrone Negus ; Thomas O'Donnell ; Dr. Markus Oldenburg ; Dr. Jaynie Padayachee ; Dave Pawson ; Nicolette Pekeur ; Sirika Pillay ; Jacques Plaut ; Andrea Prinsloo ; Joseph Raimondo ; Sanya Rajani ; Prof. Sergey Rakityansky ; Alastair Ramlakan ; Dr. Matina J. Rassias ; Dr. Jocelyn Read ; Dr. Matthew Reece ; Razvan Remsing ; Laura Richter ; Max Richter ; Sean Riddle ; Jonathan Reader ; Dr. David Roberts ; Evan Robinson ; Raoul Rontsch ; Dr. Andrew Rose ; Katie Ross ; Jeanne-Marié Roux ; Bianca Ruddy ; Katie Russell ; Steven Sam ; Nathaniel Schwartz ; Duncan Scott ; Helen Seals ; Ian Sherratt ; Dr. James Short ; Roger Sieloff ; Clare Slotow ; Bradley Smith ; Greg Solomon ; Dr. Andrew Stacey ; Dr. Jim Stasheff ; Mike Stay ; Mike Stringer ; Tim Teatro ; Ben Thompson ; Shen Tian ; Nicola du Toit ; Robert Torregrosa ; Jimmy Tseng ; Pieter Vergeer ; Helen Waugh ; Dr. Dawn Webber ; Michelle Wen ; Neels van der Westhuizen ; Dr. Alexander Wetzler ; Dr. Spencer Wheaton ; Vivian White ; Dr. Gerald Wigger ; Harry Wiggins ; Heather Williams ; Wendy Williams ; Julie Wilson ; Timothy Wilson ; Andrew Wood ; Emma Wormauld ; Dr. Sahal Yacoob ; Jean Youssef ; Ewald Zietsman

Contributors and editors have made a sincere effort to produce an accurate and useful resource.

Should you have suggestions, find mistakes or be prepared to donate material for inclusion, please don't hesitate to contact us. We intend to work with all who are willing to help make this a continuously evolving resource!

www.fhsst.org

Contents

I	Chemistry	1
1	Organic Molecules - Grade 12	3
1.1	What is organic chemistry?	3
1.2	Sources of carbon	3
1.3	Unique properties of carbon	4
1.4	Representing organic compounds	4
1.4.1	Molecular formula	4
1.4.2	Structural formula	5
1.4.3	Condensed structural formula	5
1.5	Isomerism in organic compounds	6
1.6	Functional groups	7
1.7	The Hydrocarbons	7
1.7.1	The Alkanes	10
1.7.2	Naming the alkanes	11
1.7.3	Properties of the alkanes	15
1.7.4	Reactions of the alkanes	16
1.7.5	The alkenes	18
1.7.6	Naming the alkenes	19
1.7.7	The properties of the alkenes	21
1.7.8	Reactions of the alkenes	21
1.7.9	The Alkynes	23
1.7.10	Naming the alkynes	23
1.8	The Alcohols	24
1.8.1	Naming the alcohols	25
1.8.2	Physical and chemical properties of the alcohols	27
1.9	Carboxylic Acids	28
1.9.1	Physical Properties	29
1.9.2	Derivatives of carboxylic acids: The esters	30
1.10	The Amino Group	30
1.11	The Carbonyl Group	30
1.12	Summary	31

2	Organic Macromolecules - Grade 12	37
2.1	Polymers	37
2.2	How do polymers form?	38
2.2.1	Addition polymerisation	38
2.2.2	Condensation polymerisation	40
2.3	The chemical properties of polymers	42
2.4	Types of polymers	43
2.5	Plastics	43
2.5.1	The uses of plastics	44
2.5.2	Thermoplastics and thermosetting plastics	46
2.5.3	Plastics and the environment	47
2.6	Biological Macromolecules	48
2.6.1	Carbohydrates	49
2.6.2	Proteins	51
2.6.3	Nucleic Acids	54
2.7	Summary	56
3	Reaction Rates - Grade 12	61
3.1	Introduction	61
3.2	Factors affecting reaction rates	63
3.3	Reaction rates and collision theory	67
3.4	Measuring Rates of Reaction	69
3.5	Mechanism of reaction and catalysis	71
3.6	Chemical equilibrium	74
3.6.1	Open and closed systems	76
3.6.2	Reversible reactions	76
3.6.3	Chemical equilibrium	77
3.7	The equilibrium constant	78
3.7.1	Calculating the equilibrium constant	79
3.7.2	The meaning of K_c values	80
3.8	Le Chatelier's principle	84
3.8.1	The effect of concentration on equilibrium	84
3.8.2	The effect of temperature on equilibrium	85
3.8.3	The effect of pressure on equilibrium	86
3.9	Industrial applications	90
3.10	Summary	90
4	Electrochemical Reactions - Grade 12	93
4.1	Introduction	93
4.2	The Galvanic Cell	94
4.2.1	Half-cell reactions in the Zn-Cu cell	95
4.2.2	Components of the Zn-Cu cell	96

4.2.3	The Galvanic cell	97
4.2.4	Uses and applications of the galvanic cell	98
4.3	The Electrolytic cell	99
4.3.1	The electrolysis of copper sulphate	101
4.3.2	The electrolysis of water	102
4.3.3	A comparison of galvanic and electrolytic cells	102
4.4	Standard Electrode Potentials	103
4.4.1	The different reactivities of metals	103
4.4.2	Equilibrium reactions in half cells	103
4.4.3	Measuring electrode potential	104
4.4.4	The standard hydrogen electrode	105
4.4.5	Standard electrode potentials	107
4.4.6	Combining half cells	112
4.4.7	Uses of standard electrode potential	113
4.5	Balancing redox reactions	117
4.6	Applications of electrochemistry	122
4.6.1	Electroplating	122
4.6.2	The production of chlorine	122
4.6.3	Extraction of aluminium	124
4.7	Summary	124
5	The Chemical Industry - Grade 12	129
5.1	Introduction	129
5.2	Sasol	129
5.2.1	Sasol today: Technology and production	130
5.2.2	Sasol and the environment	134
5.3	The Chloralkali Industry	136
5.3.1	The Industrial Production of Chlorine and Sodium Hydroxide	136
5.3.2	Soaps and Detergents	140
5.4	The Fertiliser Industry	144
5.4.1	The value of nutrients	144
5.4.2	The Role of fertilisers	144
5.4.3	The Industrial Production of Fertilisers	145
5.4.4	Fertilisers and the Environment: Eutrophication	148
5.5	Electrochemistry and batteries	150
5.5.1	How batteries work	150
5.5.2	Battery capacity and energy	151
5.5.3	Lead-acid batteries	151
5.5.4	The zinc-carbon dry cell	153
5.5.5	Environmental considerations	154
5.6	Summary	155

II	Physics	161
6	Motion in Two Dimensions - Grade 12	163
6.1	Introduction	163
6.2	Vertical Projectile Motion	163
6.2.1	Motion in a Gravitational Field	163
6.2.2	Equations of Motion	165
6.2.3	Graphs of Vertical Projectile Motion	168
6.3	Conservation of Momentum in Two Dimensions	176
6.4	Types of Collisions	182
6.4.1	Elastic Collisions	182
6.4.2	Inelastic Collisions	187
6.5	Frames of Reference	192
6.5.1	Introduction	192
6.5.2	What is a <i>frame of reference</i> ?	193
6.5.3	Why are frames of reference important?	193
6.5.4	Relative Velocity	193
6.6	Summary	196
6.7	End of chapter exercises	196
7	Mechanical Properties of Matter - Grade 12	205
7.1	Introduction	205
7.2	Deformation of materials	205
7.2.1	Hooke's Law	205
7.2.2	Deviation from Hooke's Law	208
7.3	Elasticity, plasticity, fracture, creep	210
7.3.1	Elasticity and plasticity	210
7.3.2	Fracture, creep and fatigue	211
7.4	Failure and strength of materials	211
7.4.1	The properties of matter	211
7.4.2	Structure and failure of materials	212
7.4.3	Controlling the properties of materials	212
7.4.4	Steps of Roman Swordsmithing	213
7.5	Summary	214
7.6	End of chapter exercise	214
8	Work, Energy and Power - Grade 12	217
8.1	Introduction	217
8.2	Work	217
8.3	Energy	223
8.3.1	External and Internal Forces	223
8.3.2	Capacity to do Work	224
8.4	Power	230

8.5	Important Equations and Quantities	234
8.6	End of Chapter Exercises	235
9	Doppler Effect - Grade 12	237
9.1	Introduction	237
9.2	The Doppler Effect with Sound and Ultrasound	237
9.2.1	Ultrasound and the Doppler Effect	241
9.3	The Doppler Effect with Light	242
9.3.1	The Expanding Universe	242
9.4	Summary	243
9.5	End of Chapter Exercises	243
10	Colour - Grade 12	245
10.1	Introduction	245
10.2	Colour and Light	245
10.2.1	Dispersion of white light	248
10.3	Addition and Subtraction of Light	248
10.3.1	Additive Primary Colours	248
10.3.2	Subtractive Primary Colours	249
10.3.3	Complementary Colours	250
10.3.4	Perception of Colour	250
10.3.5	Colours on a Television Screen	251
10.4	Pigments and Paints	252
10.4.1	Colour of opaque objects	252
10.4.2	Colour of transparent objects	252
10.4.3	Pigment primary colours	253
10.5	End of Chapter Exercises	254
11	2D and 3D Wavefronts - Grade 12	257
11.1	Introduction	257
11.2	Wavefronts	257
11.3	The Huygens Principle	258
11.4	Interference	260
11.5	Diffraction	261
11.5.1	Diffraction through a Slit	263
11.6	Shock Waves and Sonic Booms	267
11.6.1	Subsonic Flight	267
11.6.2	Supersonic Flight	268
11.6.3	Mach Cone	270
11.7	End of Chapter Exercises	272

12 Wave Nature of Matter - Grade 12	275
12.1 Introduction	275
12.2 de Broglie Wavelength	275
12.3 The Electron Microscope	278
12.3.1 Disadvantages of an Electron Microscope	280
12.3.2 Uses of Electron Microscopes	281
12.4 End of Chapter Exercises	281
13 Electrodynamics - Grade 12	283
13.1 Introduction	283
13.2 Electrical machines - generators and motors	283
13.2.1 Electrical generators	284
13.2.2 Electric motors	286
13.2.3 Real-life applications	288
13.2.4 Exercise - generators and motors	289
13.3 Alternating Current	289
13.3.1 Exercise - alternating current	291
13.4 Capacitance and inductance	291
13.4.1 Capacitance	291
13.4.2 Inductance	291
13.4.3 Exercise - capacitance and inductance	293
13.5 Summary	294
13.6 End of chapter exercise	294
14 Electronics - Grade 12	295
14.1 Introduction	295
14.2 Capacitive and Inductive Circuits	295
14.3 Filters and Signal Tuning	300
14.3.1 Capacitors and Inductors as Filters	300
14.3.2 LRC Circuits, Resonance and Signal Tuning	301
14.4 Active Circuit Elements	303
14.4.1 The Diode	303
14.4.2 The Light-Emitting Diode (LED)	305
14.4.3 Transistor	307
14.4.4 The Operational Amplifier	311
14.5 The Principles of Digital Electronics	313
14.5.1 Logic Gates	314
14.6 Using and Storing Binary Numbers	320
14.6.1 Binary numbers	320
14.6.2 Counting circuits	321
14.6.3 Storing binary numbers	323

15 EM Radiation	327
15.1 Introduction	327
15.2 Particle/wave nature of electromagnetic radiation	327
15.3 The wave nature of electromagnetic radiation	328
15.4 Electromagnetic spectrum	328
15.5 The particle nature of electromagnetic radiation	331
15.5.1 Exercise - particle nature of EM waves	332
15.6 Penetrating ability of electromagnetic radiation	332
15.6.1 Ultraviolet(UV) radiation and the skin	333
15.6.2 Ultraviolet radiation and the eyes	333
15.6.3 X-rays	334
15.6.4 Gamma-rays	334
15.6.5 Exercise - Penetrating ability of EM radiation	334
15.7 Summary	335
15.8 End of chapter exercise	335
16 Optical Phenomena and Properties of Matter - Grade 12	337
16.1 Introduction	337
16.2 The transmission and scattering of light	337
16.2.1 Energy levels of an electron	337
16.2.2 Interaction of light with metals	338
16.2.3 Why is the sky blue?	339
16.3 The photoelectric effect	340
16.3.1 Applications of the photoelectric effect	342
16.3.2 Real-life applications	344
16.4 Emission and absorption spectra	345
16.4.1 Emission Spectra	345
16.4.2 Absorption spectra	346
16.4.3 Colours and energies of electromagnetic radiation	348
16.4.4 Applications of emission and absorption spectra	350
16.5 Lasers	352
16.5.1 How a laser works	354
16.5.2 A simple laser	356
16.5.3 Laser applications and safety	357
16.6 Summary	359
16.7 End of chapter exercise	359
III Exercises	361
17 Exercises	363
IV Essays	365
Essay 1: Energy and electricity. Why the fuss?	367

18 Essay: How a cell phone works	373
19 Essay: How a Physiotherapist uses the Concept of Levers	375
20 Essay: How a Pilot Uses Vectors	377
A GNU Free Documentation License	379

Part I

Chemistry

Chapter 1

Organic Molecules - Grade 12

1.1 What is organic chemistry?

Organic chemistry is the branch of chemistry that deals with **organic molecules**. An organic molecule is one which contains **carbon**, and these molecules can range in size from simple molecules to complex structures containing thousands of atoms! Although carbon is present in all organic compounds, other elements such as hydrogen (H), oxygen (O), nitrogen (N), sulfur (S) and phosphorus (P) are also common in these molecules.

Until the early nineteenth century, chemists had managed to make many simple compounds in the laboratory, but were still unable to produce the complex molecules that they found in living organisms. It was around this time that a Swedish chemist called **Jons Jakob Berzelius** suggested that compounds found only in living organisms (the organic compounds) should be grouped separately from those found in the non-living world (the inorganic compounds). He also suggested that the laws that governed how organic compounds formed, were different from those for inorganic compounds. From this, the idea developed that there was a 'vital force' in organic compounds. In other words, scientists believed that organic compounds would not follow the normal physical and chemical laws that applied to other inorganic compounds because the very 'force of life' made them different.

This idea of a mystical 'vital force' in organic compounds was weakened when scientists began to manufacture organic compounds in the laboratory from non-living materials. One of the first to do this was **Friedrich Wohler** in 1828, who successfully prepared urea, an organic compound in the urine of animals which, until that point, had only been found in animals. A few years later a student of Wohler's, **Hermann Kolbe**, made the organic compound *acetic acid* from inorganic compounds. By this stage it was acknowledged that organic compounds are governed by exactly the same laws that apply to inorganic compounds. The properties of organic compounds are not due to a 'vital force' but to the unique properties of the carbon atom itself.

Organic compounds are very important in daily life. They make up a big part of our own bodies, they are in the food we eat and in the clothes we wear. Organic compounds are also used to make products such as medicines, plastics, washing powders, dyes, along with a list of other items.

1.2 Sources of carbon

The main source of the carbon in organic compounds is **carbon dioxide** in the air. Plants use sunlight to convert carbon dioxide into organic compounds through the process of **photosynthesis**. Plants are therefore able to make their own organic compounds through photosynthesis, while animals feed on plants or plant products so that they gain the organic compounds that they need to survive.

Another important source of carbon is **fossil fuels** such as coal, petroleum and natural gas. This

is because fossil fuels are themselves formed from the decaying remains of dead organisms (refer to Grade 11 for more information on fossil fuels).

1.3 Unique properties of carbon

Carbon has a number of unique properties which influence how it behaves and how it bonds with other atoms:

- Carbon has *four valence electrons* which means that each carbon atom can form four bonds with other atoms. Because of this, long *chain structures* can form. These chains can either be *unbranched* (figure 1.1) or *branched* (figure 1.2). Because of the number of bonds that carbon can form with other atoms, organic compounds can be very complex.

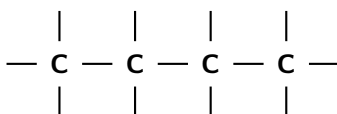


Figure 1.1: An unbranched carbon chain

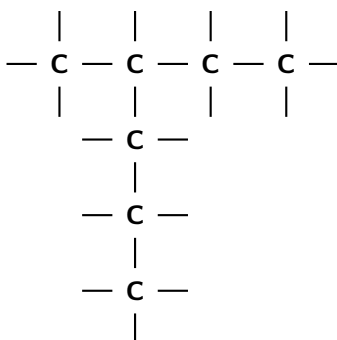


Figure 1.2: A branched carbon chain

- Because of its position on the Periodic Table, most of the bonds that carbon forms with other atoms are *covalent*. Think for example of a C-C bond. The difference in electronegativity between the two atoms is zero, so this is a pure covalent bond. In the case of a C-H bond, the difference in electronegativity between carbon (2.5) and hydrogen (2.1) is so small that C-H bonds are almost purely covalent. The result of this is that most organic compounds are non-polar. This affects some of the properties of organic compounds.

1.4 Representing organic compounds

There are a number of ways to represent organic compounds. It is useful to know all of these so that you can recognise a molecule however it is shown. There are three main ways of representing a compound. We will use the example of a molecule called 2-methylpropane to help explain the difference between each.

1.4.1 Molecular formula

The molecular formula of a compound shows how many atoms of each type are in a molecule. The number of each atom is written as a subscript after the atomic symbol. The molecular formula of 2-methylpropane is:



1.4.2 Structural formula

The structural formula of an organic compound shows every bond between every atom in the molecule. Each bond is represented by a line. The structural formula of 2-methylpropane is shown in figure 1.3.

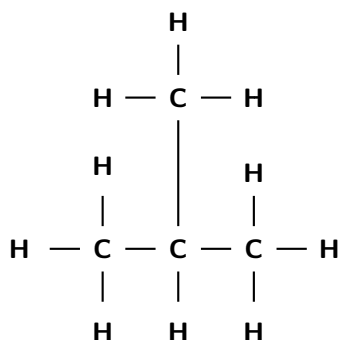
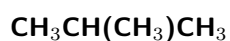


Figure 1.3: The structural formula of 2-methylpropane

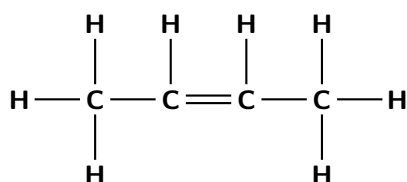
1.4.3 Condensed structural formula

When a compound is represented using its condensed structural formula, each carbon atom and the hydrogen atoms that are bonded directly to it are listed as a molecular formula, followed by a similar molecular formula for the neighbouring carbon atom. Branched groups are shown in brackets after the carbon atom to which they are bonded. The condensed structural formula below shows that in 2-methylpropane, there is a branched chain attached to the second carbon atom of the main chain. You can check this by looking at the structural formula in figure ??.

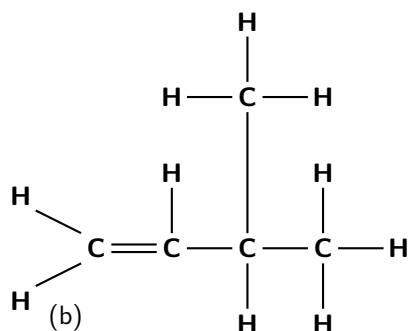


Exercise: Representing organic compounds

- For each of the following organic compounds, give the **condensed structural formula** and the **molecular formula**.



(a)



2. For each of the following, give the **structural formula** and the **molecular formula**.
- $\text{CH}_3\text{CH}_2\text{CH}_3$
 - $\text{CH}_3\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_3$
 - C_2H_6
3. Give two possible structural formulae for the compound with a molecular formula of C_4H_{10} .
-

1.5 Isomerism in organic compounds

It is possible for two organic compounds to have the same *molecular formula* but a different *structural formula*. Look for example at the two organic compounds that are shown in figure 1.4.

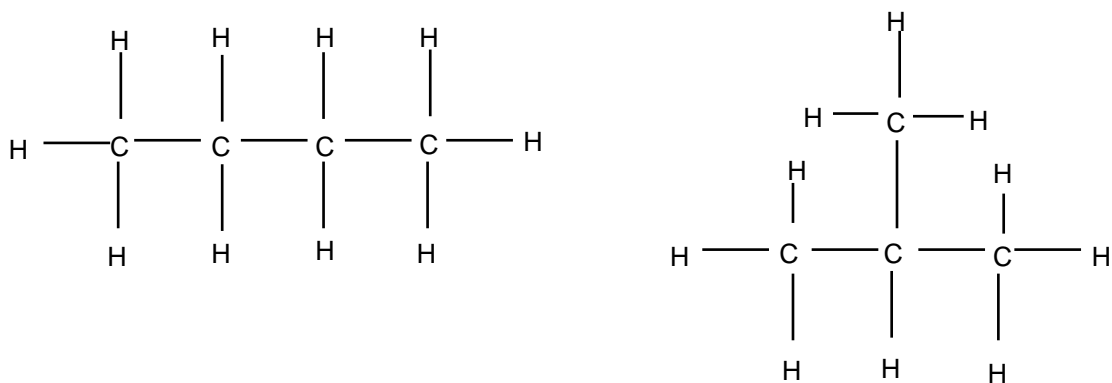


Figure 1.4: Isomers of a 4-carbon organic compound

If you were to count the number of carbon and hydrogen atoms in each compound, you would find that they are the same. They both have the same molecular formula (C_4H_{10}), but their structure is different and so are their properties. Such compounds are called **isomers**.



Definition: Isomer

In chemistry, isomers are molecules with the same molecular formula and often with the same kinds of chemical bonds between atoms, but in which the atoms are arranged differently.

**Exercise: Isomers**

Match the organic compound in Column A with its isomer Column B:

Column A	Column B
$\text{CH}_3\text{CH}(\text{CH}_3)\text{OH}$ <pre> H H H H H — C — C — C — C — H H H H H </pre>	$\text{CH}_3\text{CH}(\text{CH}_3)\text{CH}_3$ <pre> H H CH₃ H — C — C — C — H H H H </pre>
$\text{C}_3\text{H}_7\text{OH}$ <pre> CH₃ H H H — C — C — C — H H H H </pre>	$\text{C}_3\text{H}_7\text{OH}$

1.6 Functional groups

All organic compounds have a particular bond or group of atoms which we call its **functional group**. This group is important in determining how a compound will react.

**Definition: Functional group**

In organic chemistry, a functional group is a specific group of atoms within molecules, that are responsible for the characteristic chemical reactions of those molecules. The same functional group will undergo the same or similar chemical reaction(s) regardless of the size of the molecule it is a part of.

In one group of organic compounds called the **hydrocarbons**, the single, double and triple bonds of the alkanes, alkenes and alkynes are examples of functional groups. In another group, the alcohols, an oxygen and a hydrogen atom are bonded to each other to form the functional group for those compounds (in other words an alcohol has an OH in it). All alcohols will contain an oxygen and a hydrogen atom bonded together in some part of the molecule.

Table 1.1 summarises some of the common functional groups. We will look at these in more detail later in this chapter.

1.7 The Hydrocarbons

Let us first look at a group of organic compounds known as the **hydrocarbons**. These molecules only contain carbon and hydrogen. The hydrocarbons that we are going to look at are called **aliphatic compounds**. The aliphatic compounds are divided into *acyclic compounds* (chain structures) and *cyclic compounds* (ring structures). The chain structures are further divided into structures that contain only *single bonds* (**alkanes**), those that contain at least one *double bond* (**alkenes**) and those that contain at least one *triple bond* (**alkynes**). Cyclic compounds include structures such as the *benzene ring*. Figure 1.5 summarises the classification of the hydrocarbons.

Hydrocarbons that contain only single bonds are called **saturated** hydrocarbons because each carbon atom is bonded to as many hydrogen atoms as possible. Figure 1.6 shows a molecule of

Name of group	Functional group	Example	Diagram
Alkane	$\begin{array}{c} \quad \\ -C - C- \\ \quad \end{array}$	Ethane	$\begin{array}{c} H \quad H \\ \quad \\ H - C - C - H \\ \quad \\ H \quad H \end{array}$
Alkene	$\begin{array}{c} \diagdown \quad \diagup \\ C = C \\ \diagup \quad \diagdown \end{array}$	Ethene	$\begin{array}{c} H \quad \quad H \\ \diagdown \quad \diagup \\ C = C \\ \diagup \quad \diagdown \\ H \quad \quad H \end{array}$
Alkyne	$-C \equiv C-$	Ethyne (acetylene)	$H - C \equiv C - H$
Halo-alkane	$\begin{array}{c} \\ -C - X \\ \\ (X=F, Cl, Br, I) \end{array}$	Chloroethane	$\begin{array}{c} H \\ \\ CH_3 - C - X \\ \\ H \end{array}$
Alcohol/ alkanol	$\begin{array}{c} \\ -C - OH \\ \end{array}$	Ethanol	$\begin{array}{c} H \quad OH \\ \quad \\ H - C - C - H \\ \quad \\ H \quad H \end{array}$
Carboxylic acid	$\begin{array}{c} O \\ // \\ -C \\ \backslash \\ OH \end{array}$	ethanoic acid	$\begin{array}{c} O \\ \\ CH_3 - C - OH \end{array}$
Amine	$\begin{array}{c} H \\ \diagup \\ R - N \\ \diagdown \\ H \end{array}$	Glycine	$\begin{array}{c} O \quad H \quad H \\ // \quad \quad \diagup \\ C - C - N \\ \backslash \quad \quad \diagdown \\ OH \quad H \quad H \end{array}$

Table 1.1: Some functional groups of organic compounds

ethane which is a saturated hydrocarbon.

Hydrocarbons that contain double or triple bonds are called **unsaturated** hydrocarbons because

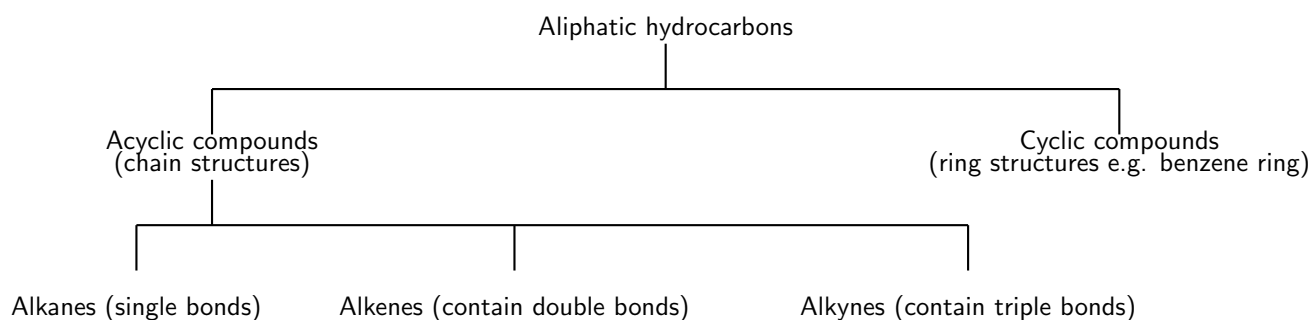


Figure 1.5: The classification of the aliphatic hydrocarbons

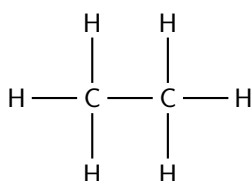


Figure 1.6: A saturated hydrocarbon

they don't contain as many hydrogen atoms as possible. Figure 1.7 shows a molecule of ethene which is an unsaturated hydrocarbon. If you compare the number of carbon and hydrogen atoms in a molecule of ethane and a molecule of ethene, you will see that the number of hydrogen atoms in ethene is *less* than the number of hydrogen atoms in ethane despite the fact that they both contain two carbon atoms. In order for an unsaturated compound to become saturated, a double bond has to be broken, and another two hydrogen atoms added for each double bond that is replaced by a single bond.

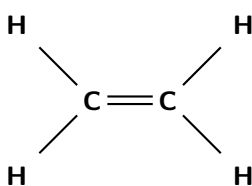
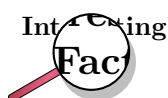


Figure 1.7: An unsaturated hydrocarbon



Fat that occurs naturally in living matter such as animals and plants is used as food for human consumption and contains varying proportions of saturated and unsaturated fat. Foods that contain a high proportion of saturated fat are butter, ghee, suet, tallow, lard, coconut oil, cottonseed oil, and palm kernel oil, dairy products (especially cream and cheese), meat, and some prepared foods. Diets high in saturated fat are correlated with an increased incidence of atherosclerosis and coronary heart disease according to a number of studies. Vegetable oils contain unsaturated fats and can be hardened to form margarine by adding hydrogen on to some of the carbon=carbon double bonds using a nickel catalyst. The process is called hydrogenation

We will now go on to look at each of the hydrocarbon groups in more detail. These groups are the alkanes, the alkenes and the alkynes.

1.7.1 The Alkanes

The alkanes are hydrocarbons that only contain *single covalent bonds* between their carbon atoms. This means that they are *saturated* compounds and are quite unreactive. The simplest alkane has only one carbon atom and is called **methane**. This molecule is shown in figure 1.8.

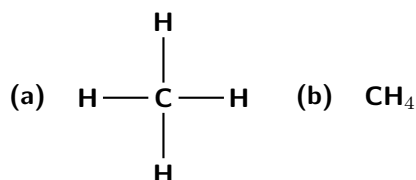


Figure 1.8: The structural (a) and molecular formula (b) for methane

The second alkane in the series has two carbon atoms and is called **ethane**. This is shown in figure 1.9.

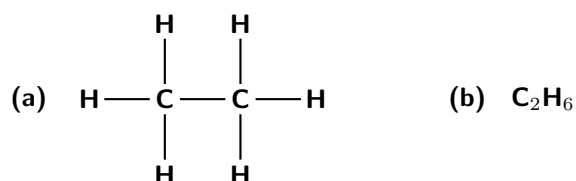


Figure 1.9: The structural (a) and molecular formula (b) for ethane

The third alkane in the series has three carbon atoms and is called **propane** (Figure 1.10).

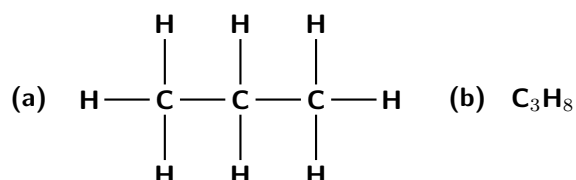
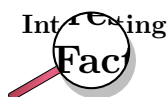


Figure 1.10: The structural (a) and molecular formula (b) for propane

When you look at the molecular formula for each of the alkanes, you should notice a pattern developing. For each carbon atom that is added to the molecule, two hydrogen atoms are added. In other words, each molecule differs from the one before it by CH_2 . This is called a *homologous series*. The alkanes have the general formula $\text{C}_n\text{H}_{2n+2}$.

The alkanes are the most important source of fuel in the world and are used extensively in the chemical industry. Some are gases (e.g. methane and ethane), while others are liquid fuels (e.g. octane, an important component of petrol).



Some fungi use alkanes as a source of carbon and energy. One fungus *Amorphotheca resinae* prefers the alkanes used in aviation fuel, and this can cause problems for aircraft in tropical areas!

1.7.2 Naming the alkanes

In order to give compounds a name, certain rules must be followed. When naming organic compounds, the IUPAC (International Union of Pure and Applied Chemistry) nomenclature is used. We will first look at some of the steps that need to be followed when naming a compound, and then try to apply these rules to some specific examples.

1. STEP 1: Recognise the *functional group* in the compound. This will determine the suffix (the 'end') of the name. For example, if the compound is an alkane, the suffix will be -ane; if the compound is an alkene the suffix will be -ene; if the compound is an alcohol the suffix will be -ol, and so on.
2. STEP 2: Find the longest continuous carbon chain (it won't always be a *straight* chain) and count the number of carbon atoms in this chain. This number will determine the prefix (the 'beginning') of the compound's name. These prefixes are shown in table 1.2. So, for example, an alkane that has 3 carbon atoms will have the suffix *prop* and the compound's name will be *propane*.

Carbon atoms	prefix
1	meth(ane)
2	eth(ane)
3	prop(ane)
4	but(ane)
5	pent(ane)
6	hex(ane)
7	hept(ane)
8	oct(ane)
9	non(ane)
10	dec(ane)

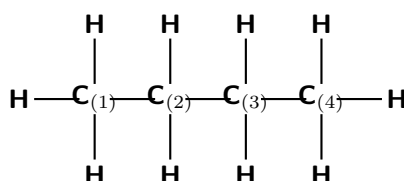
Table 1.2: The prefix of a compound's name is determined by the number of carbon atoms in the longest chain

3. STEP 3: Number the carbons in the longest carbon chain (Important: If there is a double or triple bond, you need to start numbering so that the bond is at the carbon with the lowest number).
4. STEP 4: Look for any branched groups and name them. Also give them a number to show their position on the carbon chain. If there are no branched groups, this step can be left out.
5. STEP 5: Combine the elements of the name into a single word in the following order: branched groups; prefix; name ending according to the functional group and its position along the longest carbon chain.



Worked Example 1: Naming the alkanes

Question: Give the IUPAC name for the following compound:



Note: The numbers attached to the carbon atoms would not normally be shown. The atoms have been numbered to help you to name the compound.

Answer

Step 1 : Identify the functional group

The compound is a hydrocarbon with single bonds between the carbon atoms. It is an alkane and will have a suffix of -ane.

Step 2 : Find the longest carbon chain

There are four carbon atoms in the longest chain. The prefix of the compound will be 'but'.

Step 3 : Number the carbons in the longest chain

In this case, it is easy. The carbons are numbered from left to right, from one to four.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain

There are no branched groups in this compound.

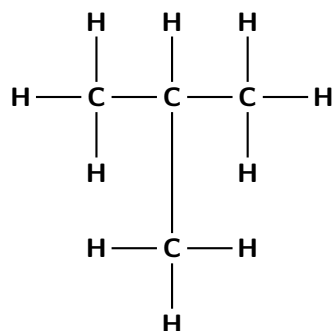
Step 5 : Combine the elements of the name into a single word

The name of the compound is **butane**.



Worked Example 2: Naming the alkanes

Question: Give the IUPAC name for the following compound:



Answer

Step 1 : Identify the functional group

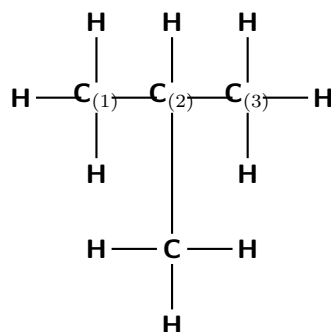
The compound is an alkane and will have the suffix -ane.

Step 2 : Find the longest carbon chain

There are three carbons in the longest chain. The prefix for this compound is -prop.

Step 3 : Number the carbons in the carbon chain

If we start at the carbon on the left, we can number the atoms as shown below:



Step 4 : Look for any branched groups, name them and give their position on the carbon chain

There is a branched group attached to the second carbon atom. This group has the formula CH_3 which is methane. However, because it is not part of the main chain, it is given the suffix -yl (i.e. methyl). The position of the methyl group comes just before its name (see next step).

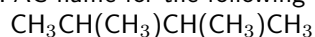
Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.

The compound's name is **2-methylpropane**.



Worked Example 3: Naming the alkanes

Question: Give the IUPAC name for the following compound:

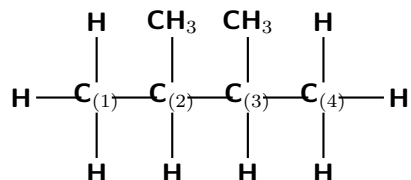


(Remember that the side groups are shown in brackets after the carbon atom to which they are attached.)

Answer

Step 1 : Draw the compound from its condensed structural formula

The structural formula of the compound is:



Step 2 : Identify the functional group

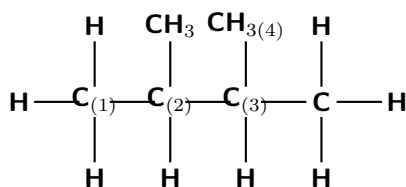
The compound is an alkane and will have the suffix -ane.

Step 3 : Find the longest carbon chain

There are four carbons in the longest chain. The prefix for this compound is -but.

Step 4 : Number the carbons in the carbon chain

If we start at the carbon on the left, carbon atoms are numbered as shown in the diagram above. A second way that the carbons could be numbered is:



Step 5 : Look for any branched groups, name them and give their position on the carbon chain

There are two methyl groups attached to the main chain. The first one is attached to the second carbon atom and the second methyl group is attached to the third carbon atom. Notice that in this example it does not matter how you have chosen to number the carbons in the main chain; the methyl groups are still attached to the second and third carbons and so the naming of the compound is not affected.

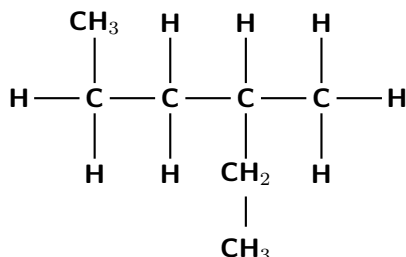
Step 6 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.

The compound's name is **2,3-dimethyl-butane**.



Worked Example 4: Naming the alkanes

Question: Give the IUPAC name for the following compound:



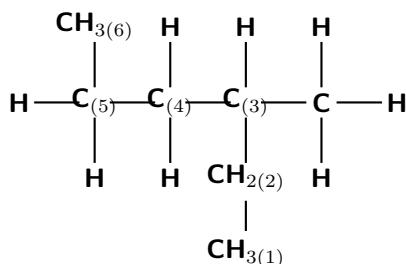
Answer

Step 1 : Identify the functional group

The compound is an alkane and will have the suffix -ane.

Step 2 : Find the longest carbon chain and number the carbons in the longest chain.

There are six carbons in the longest chain if they are numbered as shown below. The prefix for the compound is hex-.



Step 3 : Look for any branched groups, name them and give their position on the carbon chain

There is one methyl group attached to the main chain. This is attached to the third carbon atom.

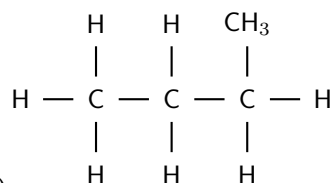
Step 4 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.

The compound's name is **3-methyl-hexane**.



Exercise: Naming the alkanes

- Give the structural formula for each of the following:
 - Octane
 - $\text{CH}_3\text{CH}_2\text{CH}_3$
 - $\text{CH}_3\text{CH}(\text{CH}_3)\text{CH}_3$
 - 3-ethyl-pentane
- Give the IUPAC name for each of the following organic compounds.



-
- $\text{CH}_3\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_2\text{CH}_3$
- $\text{CH}_3\text{CH}(\text{CH}_3)\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_3$

1.7.3 Properties of the alkanes

We have already mentioned that the alkanes are relatively unreactive because of their stable C-C and C-H bonds. The boiling point and melting point of these molecules is determined by their molecular structure, and their surface area. The more carbon atoms there are in an alkane, the greater the surface area and therefore the higher the boiling point. The melting point also increases as the number of carbon atoms in the molecule increases. This can be seen in the data in table 1.3.

Formula	Name	Melting point ($^{\circ}\text{C}$)	Boiling point ($^{\circ}\text{C}$)	Phase at room temperature
CH_4	methane	-183	-162	gas
C_2H_6	ethane	-182	-88	gas
C_3H_8	propane	-187	-45	gas
C_4H_{10}	butane	-138	-0.5	gas
C_5H_{12}	pentane	-130	36	liquid
C_6H_{14}	hexane	-95	69	liquid
$\text{C}_{17}\text{H}_{36}$	heptadecane	22	302	solid

Table 1.3: Properties of some of the alkanes

You will also notice that, when the molecular mass of the alkanes is low (i.e. there are few carbon atoms), the organic compounds are *gases* because the intermolecular forces are weak. As the number of carbon atoms and the molecular mass increases, the compounds are more likely to be liquids or solids because the intermolecular forces are stronger.

1.7.4 Reactions of the alkanes

There are three types of reactions that can occur in saturated compounds such as the alkanes.

1. Substitution reactions

Substitution reactions involve the removal of a hydrogen atom which is replaced by an atom of another element, such as a halogen (F, Cl, Br or I) (figure 1.11). The product is called a **halo-alkane**. Since alkanes are not very reactive, either heat or light is needed for this reaction to take place.

e.g. $\text{CH}_2=\text{CH}_2 + \text{HBr} \rightarrow \text{CH}_3\text{-CH}_2\text{-Br}$ (halo-alkane)

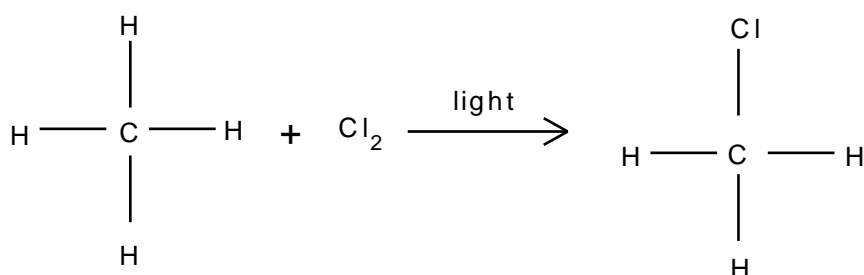


Figure 1.11: A substitution reaction

Halo-alkanes (also sometimes called *alkyl halides*) that contain methane and chlorine are substances that can be used as anaesthetics during operations. One example is trichloromethane, also known as 'chloroform' (figure 1.12).

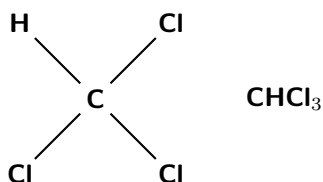


Figure 1.12: Trichloromethane

2. Elimination reactions

Saturated compounds can also undergo elimination reactions to become unsaturated (figure 1.13). In the example below, an atom of hydrogen and chlorine are eliminated from the original compound to form an unsaturated halo-alkene.

e.g. $\text{ClH}_2\text{C} - \text{CH}_2\text{Cl} \rightarrow \text{H}_2\text{C} = \text{CHCl} + \text{HCl}$

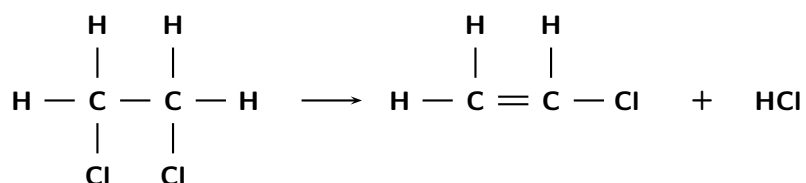


Figure 1.13: An elimination reaction

3. Oxidation reactions

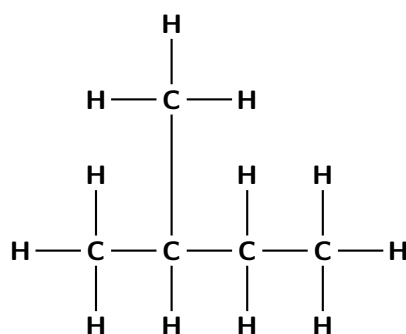
When alkanes are burnt in air, they react with the oxygen in air and heat is produced. This is called an oxidation or combustion reaction. Carbon dioxide and water are given off as products. Heat is also released during the reaction. The burning of alkanes provides most of the energy that is used by man.

e.g. $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O + \text{heat}$

**Exercise: The Alkanes**

1. Give the IUPAC name for each of the following alkanes:

(a) $CH_3CH_2CH_2CH_2CH_2CH_3$



(b)

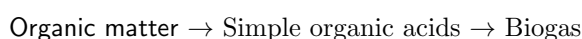
(c) CH_3CH_3

2. Give the structural formula for each of the following compounds:

(a) octane

(b) 3-methyl-hexane

3. Methane is one of the simplest alkanes and yet it is an important fuel source. Methane occurs naturally in wetlands, natural gas and permafrost. However, methane can also be produced when organic wastes (e.g. animal manure and decaying material) are broken down by bacteria under conditions that are anaerobic (there is no oxygen). The simplified reaction is shown below:



The organic matter could be carbohydrates, proteins or fats which are broken down by acid-forming bacteria into simple organic acids such as acetic acid or formic acid. Methane-forming bacteria then convert these acids into biogases such as methane and ammonia.

The production of methane in this way is very important because methane can be used as a fuel source. One of the advantages of methane over other fuels like coal, is that it produces more energy but with lower carbon dioxide emissions. The problem however, is that methane itself is a greenhouse gas and has a much higher global warming potential than carbon dioxide. So, producing methane may in fact have an even more dangerous impact on the environment.

(a) What is the structural formula of methane?

(b) Write an equation to show the reaction that takes place when methane is burned as a fuel.

(c) Explain what is meant by the statement that methane 'has a greater global warming potential than carbon dioxide'.

4. Chlorine and ethane react to form chloroethane and hydrogen chloride.

(a) Write a balanced chemical equation for this reaction, using molecular formulae.

- (b) Give the structural formula of chloroethane.
 (c) What type of reaction has taken place in this example?
5. Petrol (C_8H_{18}) is in fact not pure C_8H_{18} but a mixture of various *alkanes*. The 'octane rating' of petrol refers to the percentage of the petrol which is C_8H_{18} . For example, 93 octane fuel contains 93% C_8H_{18} and 7% other alkanes. The *isomer* of C_8H_{18} referred to in the 'octane rating' is in fact not octane but 2,2,4-trimethylpentane.
- (a) Write an unbalanced equation for the chemical reaction which takes place when petrol (C_8H_{18}) burns in excess oxygen.
 (b) Write the general formula of the *alkanes*.
 (c) Define the term *structural isomer*.
 (d) Use the information given in this question and your knowledge of naming organic compounds to deduce and draw the full structural formula for 2,2,4-trimethylpentane. (IEB pg 25)
-

1.7.5 The alkenes

In the alkenes, there is at least one double bond between two carbon atoms. This means that they are **unsaturated** and are *more reactive* than the alkanes. The simplest alkene is ethene (also known as ethylene), which is shown in figure 1.14.

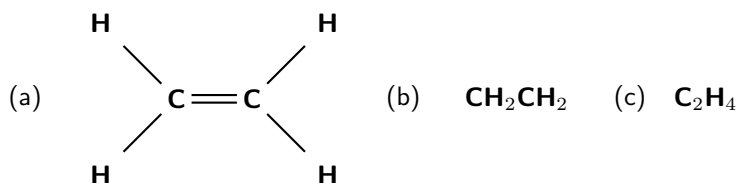


Figure 1.14: The (a) structural, (b) condensed structural and (c) molecular structure representations of ethene

As with the alkanes, the alkenes also form a homologous series. They have the general formula C_nH_{2n} . The second alkene in the series would therefore be C_3H_6 . This molecule is known as propene (figure 1.15). Note that if an alkene has two double bonds, it is called a **diene** and if it has three double bonds it is called a **triene**.

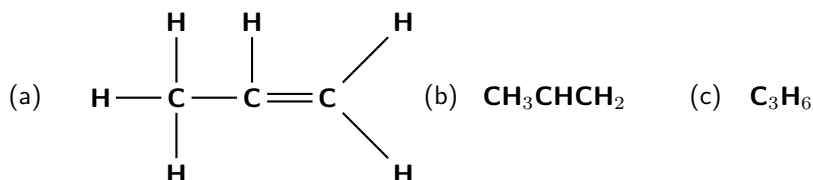


Figure 1.15: The (a) structural, (b) condensed structural and (c) molecular structure representations of propene

The alkenes have a variety of uses. Ethylene for example is a hormone in plants that stimulates the ripening of fruits and the opening of flowers. Propene is an important compound in the petrochemicals industry. It is used as a monomer to make polypropylene and is also used as a fuel gas for other industrial processes.

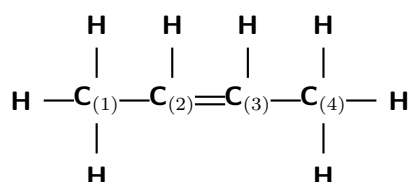
1.7.6 Naming the alkenes

Similar rules will apply in naming the alkenes, as for the alkanes.



Worked Example 5: Naming the alkenes

Question: Give the IUPAC name for the following compound:



Answer

Step 1 : Identify the functional group

The compound is an alkene and will have the suffix -ene.

Step 2 : Find the longest carbon chain

There are four carbon atoms in the longest chain and so the prefix for this compound will be 'but'.

Step 3 : Number the carbon atoms

Remember that when there is a double or triple bond, the carbon atoms must be numbered so that the double or triple bond is at the lowest numbered carbon. In this case, it doesn't matter whether we number the carbons from the left to right, or from the right to left. The double bond will still fall between C₂ and C₃. The position of the bond will come just before the suffix in the compound's name.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain

There are no branched groups in this molecule.

Step 5 : Name the compound

The name of this compound is **but-2-ene**.



Worked Example 6: Naming the alkenes

Question: Draw the structural formula for the organic compound 3-methylbutene

Answer

Step 6 : Identify the functional group

The suffix -ene means that this compound is an alkene and there must be a double bond in the molecule. There is no number immediately before the suffix which means that the double bond must be at the first carbon in the chain.

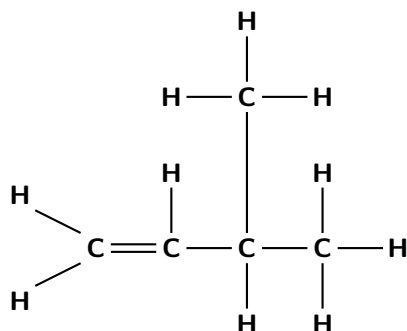
Step 7 : Determine the number of carbons in the longest chain

The prefix for the compound is 'but' so there must be four carbons in the longest chain.

Step 8 : Look for any branched groups

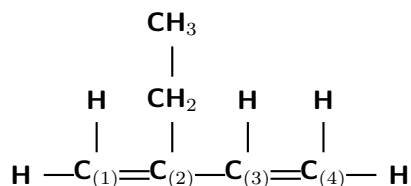
There is a methyl group at the third carbon atom in the chain.

Step 9 : Combine this information to draw the structural formula for this molecule.



Worked Example 7: Naming the alkenes

Question: Give the IUPAC name for the following compound:



Answer

Step 1 : Identify the functional group

The compound is an alkene and will have the suffix -ene. There is a double bond between the first and second carbons and also between the third and fourth carbons. The organic compound is therefore a 'diene'.

Step 2 : Find the longest carbon chain and number the carbon atoms

There are four carbon atoms in the longest chain and so the prefix for this compound will be 'but'. The carbon atoms are numbered 1 to 4 in the diagram above. Remember that the main carbon chain must contain both the double bonds.

Step 3 : Look for any branched groups, name them and give their position on the carbon chain

There is an ethyl group on the second carbon.

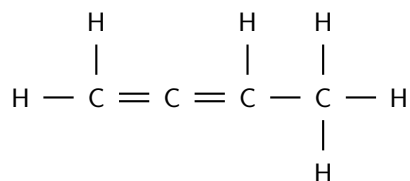
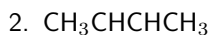
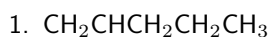
Step 4 : Name the compound

The name of this compound is **2-ethyl-but-1,3-diene**.



Exercise: Naming the alkenes

Give the IUPAC name for each of the following alkenes:



3.

1.7.7 The properties of the alkenes

The properties of the alkenes are very similar to those of the alkanes, except that the alkenes are more reactive because they are unsaturated. As with the alkanes, compounds that have four or less carbon atoms are gases at room temperature, while those with five or more carbon atoms are liquids.

1.7.8 Reactions of the alkenes

Alkenes can undergo **addition reactions** because they are unsaturated. They readily react with hydrogen, water and the halogens. The double bond is broken and a single, saturated bond is formed. A new group is then added to one or both of the carbon atoms that previously made up the double bond. The following are some examples:

1. Hydrogenation reactions

A catalyst such as platinum is normally needed for these reactions

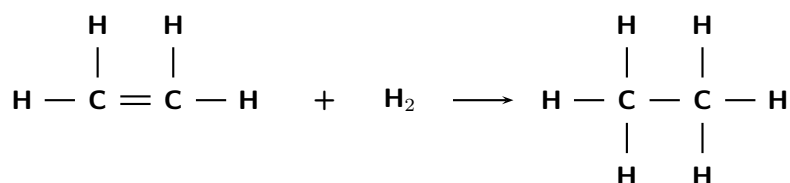
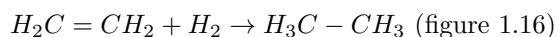


Figure 1.16: A hydrogenation reaction

2. Halogenation reactions

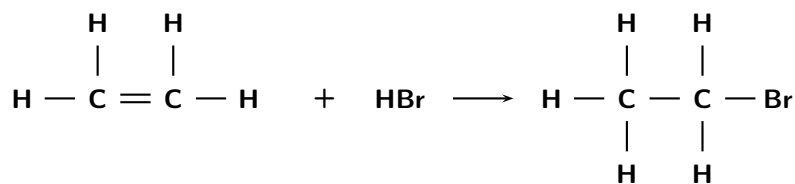
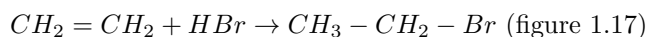
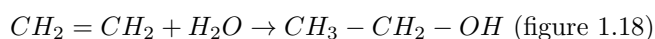


Figure 1.17: A halogenation reaction

3. The formation of alcohols



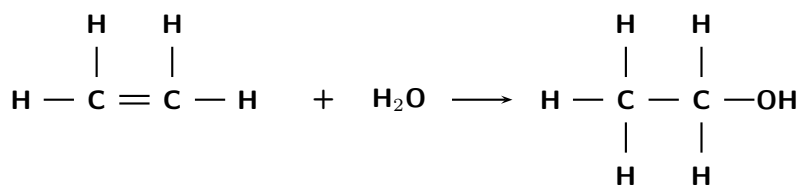
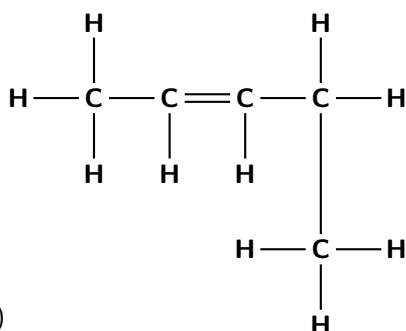


Figure 1.18: The formation of an alcohol

**Exercise: The Alkenes**

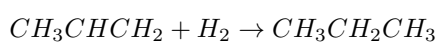
1. Give the IUPAC name for each of the following organic compounds:



- (a)
 (b) CH_3CHCH_2
2. Refer to the data table below which shows the melting point and boiling point for a number of different organic compounds.

Formula	Name	Melting point ($^{\circ}\text{C}$)	Boiling point ($^{\circ}\text{C}$)
C_4H_{10}	Butane	-138	-0.5
C_5H_{12}	Pentane	-130	36
C_6H_{14}	Hexane	-95	69
C_4H_8	Butene	-185	-6
C_5H_{10}	Pentene	-138	30
C_6H_{12}	Hexene	-140	63

- (a) At room temperature (approx. 25°C), which of the organic compounds in the table are:
- gases
 - liquids
- (b) In the alkanes...
- Describe what happens to the melting point and boiling point as the number of carbon atoms in the compound increases.
 - Explain why this is the case.
- (c) If you look at an alkane and an alkene that have the same number of carbon atoms...
- How do their melting points and boiling points compare?
 - Can you explain why their melting points and boiling points are different?
- (d) Which of the compounds, hexane or hexene, is more reactive? Explain your answer.
3. The following reaction takes place:



- (a) Give the name of the organic compound in the reactants.

- (b) What is the name of the product?
 (c) What type of reaction is this?
 (d) Which compound in the reaction is a saturated hydrocarbon?
-

1.7.9 The Alkynes

In the alkynes, there is at least one triple bond between two of the carbon atoms. They are unsaturated compounds and are therefore highly reactive. Their general formula is C_nH_{2n-2} . The simplest alkyne is ethyne (figure 1.19), also known as acetylene. Many of the alkynes are used to synthesise other chemical products.

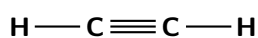
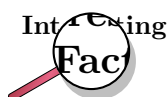
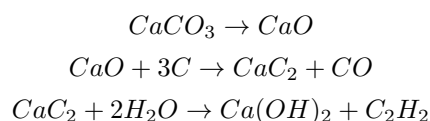


Figure 1.19: Ethyne (acetylene)



The raw materials that are needed to make acetylene are calcium carbonate and coal. Acetylene can be produced through the following reactions:



An important use of acetylene is in oxyacetylene gas welding. The fuel gas burns with oxygen in a torch. An incredibly high heat is produced, and this is enough to melt metal.

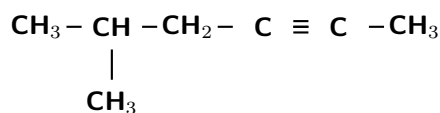
1.7.10 Naming the alkynes

The same rules will apply as for the alkanes and alkenes, except that the suffix of the name will now be -yne.



Worked Example 8: Naming the alkynes

Question: Give the IUPAC name for the following compound:



Answer**Step 5 : Identify the functional group**

There is a triple bond between two of the carbon atoms, so this compound is an alkyne. The suffix will be -yne. The triple bond is at the second carbon, so the suffix will in fact be 2-yne.

Step 6 : Find the longest carbon chain and give the compound the correct prefix

If we count the carbons in a straight line, there are six. The prefix of the compound's name will be 'hex'.

Step 7 : Number the carbons in the longest chain

In this example, you will need to number the carbons from right to left so that the triple bond is between carbon atoms with the lowest numbers.

Step 8 : Look for any branched groups, name them and show the number of the carbon atom to which the group is attached

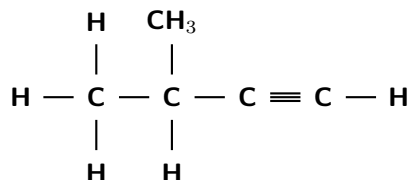
There is a methyl (CH_3) group attached to the fifth carbon (remember we have numbered the carbon atoms from right to left).

Step 9 : Combine the elements of the name into a single word in the following order: branched groups; prefix; name ending according to the functional group and its position along the longest carbon chain.

If we follow this order, the name of the compound is **5-methyl-hex-2-yne**.

**Exercise: The alkynes**

Give the IUPAC name for each of the following organic compounds.



- 1.
2. C_2H_2
3. $\text{CH}_3\text{CH}_2\text{CCH}$

1.8 The Alcohols

An alcohol is any organic compound where there is a *hydroxyl* functional group (-OH) bound to a carbon atom. The general formula for a simple alcohol is $\text{C}_n\text{H}_{2n+1}\text{OH}$.

The simplest and most commonly used alcohols are methanol and ethanol (figure 1.20).

The alcohols have a number of different uses:

- methylated spirits (surgical spirits) is a form of ethanol where methanol has been added
- ethanol is used in alcoholic drinks

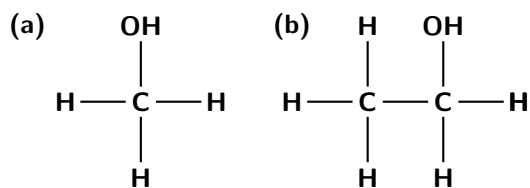
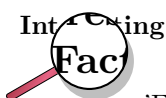
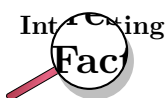
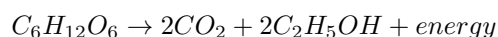


Figure 1.20: (a) methanol and (b) ethanol

- ethanol is used as an industrial solvent
- methanol and ethanol can both be used as a fuel and they burn more cleanly than gasoline or diesel (refer to Grade 11 for more information on biofuels as an alternative energy resource.)
- ethanol is used as a solvent in medical drugs, perfumes and vegetable essences
- ethanol is an antiseptic



'Fermentation' refers to the conversion of sugar to alcohol using yeast (a fungus). The process of fermentation produces items such as wine, beer and yoghurt. To make wine, grape juice is fermented to produce alcohol. This reaction is shown below:



Ethanol is a diuretic. In humans, ethanol reduces the secretion of a hormone called antidiuretic hormone (ADH). The role of ADH is to control the amount of water that the body retains. When this hormone is not secreted in the right quantities, it can cause dehydration because too much water is lost from the body in the urine. This is why people who drink too much alcohol can become dehydrated, and experience symptoms such as headaches, dry mouth, and lethargy. Part of the reason for the headaches is that dehydration causes the brain to shrink away from the skull slightly. The effects of drinking too much alcohol can be reduced by drinking lots of water.

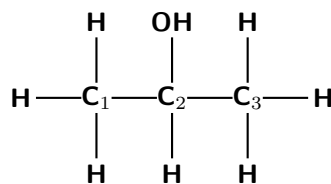
1.8.1 Naming the alcohols

The rules used to name the alcohols are similar to those already discussed for the other compounds, except that the suffix of the name will be different because the compound is an alcohol.



Worked Example 9: Naming alcohols 1

Question: Give the IUPAC name for the following organic compound

**Answer****Step 1 : Identify the functional group**

The compound has an -OH (hydroxyl) functional group and is therefore an alcohol. The compound will have the suffix -ol.

Step 2 : Find the longest carbon chain

There are three carbons in the longest chain. The prefix for this compound will be 'prop'. Since there are only single bonds between the carbon atoms, the suffix becomes 'propan' (similar to the alkane 'propane').

Step 3 : Number the carbons in the carbon chain

In this case, it doesn't matter whether you start numbering from the left or right. The hydroxyl group will still be attached to the middle carbon atom, numbered '2'.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain.

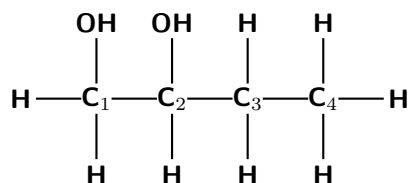
There are no branched groups in this compound, but you still need to indicate the position of the hydroxyl group on the second carbon. The suffix will be -2-ol because the hydroxyl group is attached to the second carbon.

Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.

The compound's name is propan-2-ol.

**Worked Example 10: Naming alcohols 2**

Question: Give the IUPAC name for the following compound:

**Answer****Step 1 : Identify the functional group**

The compound has an -OH (hydroxyl) functional group and is therefore an alcohol. There are two hydroxyl groups in the compound, so the suffix will be -diol.

Step 2 : Find the longest carbon chain

There are four carbons in the longest chain. The prefix for this compound will be 'butan'.

Step 3 : Number the carbons in the carbon chain

The carbons will be numbered from left to right so that the two hydroxyl groups are attached to carbon atoms with the lowest numbers.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain.

There are no branched groups in this compound, but you still need to indicate the position of the hydroxyl groups on the first and second carbon atoms. The suffix will therefore become 1,2-diol.

Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.

The compound's name is butan-1,2-diol.

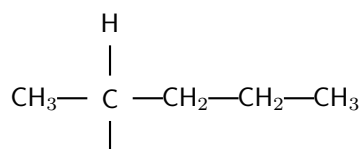
**Exercise: Naming the alcohols**

1. Give the structural formula of each of the following organic compounds:

- (a) pentan-3-ol
- (b) butan-2,3-diol
- (c) 2-methyl-propanol

2. Give the IUPAC name for each of the following:

- (a) $\text{CH}_3\text{CH}_2\text{CH}(\text{OH})\text{CH}_3$



- (b)

1.8.2 Physical and chemical properties of the alcohols

The hydroxyl group affects the **solubility** of the alcohols. The hydroxyl group generally makes the alcohol molecule *polar* and therefore more likely to be **soluble** in water. However, the carbon chain resists solubility, so there are two opposing trends in the alcohols. Alcohols with shorter carbon chains are usually more soluble than those with longer carbon chains.

Alcohols tend to have higher **boiling points** than the hydrocarbons because of the strong hydrogen bond between hydrogen and oxygen atoms.

Alcohols can show either **acidic** or **basic** properties because of the hydroxyl group. They also undergo oxidation reactions to form aldehydes, ketones and carboxylic acids.

Activity :: Case Study : The uses of the alcohols

Read the extract below and then answer the questions that follow:

The alcohols are a very important group of organic compounds, and they have a variety of uses. Our most common use of the word 'alcohol' is with reference to alcoholic drinks. The alcohol in drinks is in fact **ethanol**. But ethanol has many more uses apart from alcoholic drinks! When ethanol burns in air, it produces carbon dioxide, water and energy and can therefore be used as a fuel on its own, or in mixtures with petrol. Because ethanol can be produced through fermentation, this is a useful way for countries without an oil industry to reduce imports of petrol. Ethanol is also used as a solvent in many perfumes and cosmetics.

Methanol can also be used as a fuel, or as a petrol additive to improve combustion. Most methanol is used as an industrial feedstock, in other words it is used to make other things such as methanal (formaldehyde), ethanoic acid and methyl esters. In most cases, these are then turned into other products.

Propan-2-ol is another important alcohol, which is used in a variety of applications as a solvent.

Questions

1. Give the structural formula for propan-2-ol.
 2. Write a balanced chemical equation for the combustion reaction of ethanol.
 3. Explain why the alcohols are good solvents.
-

1.9 Carboxylic Acids

Carboxylic acids are organic acids that are characterised by having a carboxyl group, which has the formula $-(C=O)-OH$, or more commonly written as $-COOH$. In a carboxyl group, an oxygen atom is double-bonded to a carbon atom, which is also bonded to a hydroxyl group. The simplest carboxylic acid, methanoic acid, is shown in figure 1.21. The IUPAC suffix for carboxylic acids is -anoic acid.

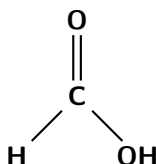
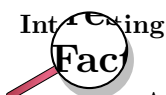


Figure 1.21: Methanoic acid

Carboxylic acids are widespread in nature. Methanoic acid (also known as *formic acid*) has the formula $HCOOH$ and is found in insect stings. Ethanoic acid (CH_3COOH), or *acetic acid*, is the main component of vinegar. More complex organic acids also have a variety of different functions. Benzoic acid (C_6H_5COOH) for example, is used as a food preservative.



A certain type of ant, called formicine ants, manufacture and secrete formic acid, which is used to defend themselves against other organisms that might try to eat them.

1.9.1 Physical Properties

Carboxylic acids are **weak acids**, in other words they only dissociate partially. Why does the carboxyl group have acidic properties? In the carboxyl group, the hydrogen tends to separate itself from the oxygen atom. In other words, the carboxyl group becomes a source of positively-charged hydrogen ions (H^+). This is shown in figure 1.22.

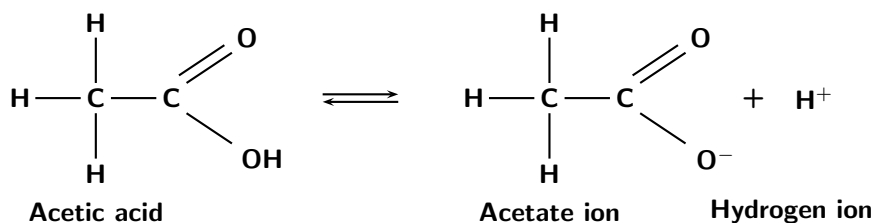


Figure 1.22: The dissociation of a carboxylic acid



Exercise: Carboxylic acids

1. Refer to the table below which gives information about a number of carboxylic acids, and then answer the questions that follow.

Formula	Common name	Source	IUPAC name	melting point ($^{\circ}\text{C}$)	boiling point ($^{\circ}\text{C}$)
	formic acid	ants	methanoic acid	8.4	101
$\text{CH}_3\text{CO}_2\text{H}$		vinegar	ethanoic acid	16.6	118
	propionic acid	milk	propanoic acid	-20.8	141
$\text{CH}_3(\text{CH}_2)_2\text{CO}_2\text{H}$	butyric acid	butter		-5.5	164
	valeric acid	valerian root	pentanoic acid	-34.5	186
$\text{CH}_3(\text{CH}_2)_4\text{CO}_2\text{H}$	caproic acid	goats		-4	205
	enanthic acid	vines		-7.5	223
$\text{CH}_3(\text{CH}_2)_6\text{CO}_2\text{H}$	caprylic acid	goats		16.3	239

- (a) Fill in the missing spaces in the table by writing the formula, common name or IUPAC name.
- (b) Draw the structural formula for butyric acid.
- (c) Give the molecular formula for caprylic acid.
- (d) Draw a graph to show the relationship between molecular mass (on the x-axis) and boiling point (on the y-axis)
 - i. Describe the trend you see.
 - ii. Suggest a reason for this trend.

1.9.2 Derivatives of carboxylic acids: The esters

When an alcohol reacts with a carboxylic acid, an **ester** is formed. Most esters have a characteristic and pleasant smell. In the reaction, the hydrogen atom from the hydroxyl group, and an OH from the carboxylic acid, form a molecule of water. A new bond is formed between what remains of the alcohol and acid. The name of the ester is a combination of the names of the alcohol and carboxylic acid. The suffix for an ester is -oate. An example is shown in figure 1.23.

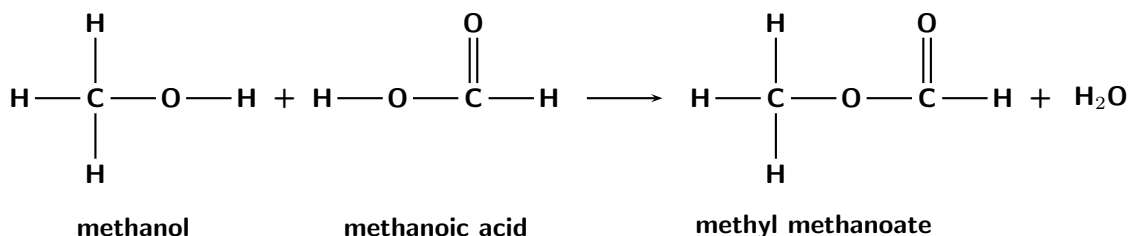


Figure 1.23: The formation of an ester and water from an alcohol and carboxylic acid

1.10 The Amino Group

The amino group has the formula $-\text{NH}_2$ and consists of a nitrogen atom that is bonded to two hydrogen atoms, and to the carbon skeleton. Organic compounds that contain this functional group are called **amines**. One example is *glycine*. Glycine belongs to a group of organic compounds called *amino acids*, which are the building blocks of proteins.

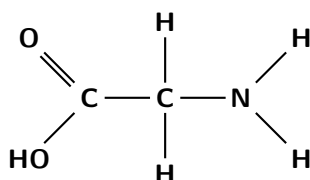


Figure 1.24: A molecule of glycine

1.11 The Carbonyl Group

The carbonyl group ($-\text{CO}$) consists of a carbon atom that is joined to an oxygen by a double bond. If the functional group is on the *end* of the carbon chain, the organic compound is called a **ketone**. The simplest ketone is *acetone*, which contains three carbon atoms. A ketone has the ending 'one' in its IUPAC name.



Exercise: Carboxylic acids, esters, amines and ketones

- Look at the list of organic compounds in the table below:

Organic compound	Type of compound
$\text{CH}_3\text{CH}_2\text{CH}_2\text{COOH}$	
$\text{NH}_2\text{CH}_2\text{COOH}$	
propyl ethanoate	
CH_3CHO	

- (a) Complete the table by identifying each compound as either a carboxylic acid, ester, amine or ketone.
 - (b) Give the name of the compounds that have been written as condensed structural formulae.
2. A chemical reaction takes place and ethyl methanoate is formed.
- (a) What type of organic compound is ethyl methanoate?
 - (b) Name the two reactants in this chemical reaction.
 - (c) Give the structural formula of ethyl methanoate.
-

1.12 Summary

- **Organic chemistry** is the branch of chemistry that deals with organic molecules. An **organic molecule** is one that contains carbon.
- All **living organisms** contain carbon. Plants use sunlight to convert carbon dioxide in the air into organic compounds through the process of **photosynthesis**. Animals and other organisms then feed on plants to obtain their own organic compounds. **Fossil fuels** are another important source of carbon.
- It is the **unique properties of the carbon atom** that give organic compounds certain properties.
- The carbon atom has **four valence electrons**, so it can bond with many other atoms, often resulting in long chain structures. It also forms mostly **covalent bonds** with the atoms that it bonds to, meaning that most organic molecules are **non-polar**.
- An organic compound can be represented in different ways, using its **molecular formula**, **structural formula** or **condensed structural formula**.
- If two compounds are **isomers**, it means that they have the same molecular formulae but different structural formulae.
- A **functional group** is a particular group of atoms within a molecule, which give it certain reaction characteristics. Organic compounds can be grouped according to their functional group.
- The **hydrocarbons** are organic compounds that contain only carbon and hydrogen. They can be further divided into the alkanes, alkenes and alkynes, based on the type of bonds between the carbon atoms.
- The **alkanes** have only **single bonds** between their carbon atoms and are unreactive.
- The **alkenes** have at least one **double bond** between two of their carbon atoms. They are more reactive than the alkanes.
- The **alkynes** have at least one **triple bond** between two of their carbon atoms. They are the most reactive of the three groups.
- A hydrocarbon is said to be **saturated** if it contains the maximum possible number of hydrogen atoms for that molecule. The alkanes are all saturated compounds.
- A hydrocarbon is **unsaturated** if it does not contain the maximum number of hydrogen atoms for that molecule. The alkenes and alkynes are examples of unsaturated molecules. If a double or triple bond is broken, more hydrogen atoms can be added to the molecule.
- There are three types of reactions that occur in the alkanes: **substitution**, **elimination** and **oxidation** reactions.
- The alkenes undergo **addition** reactions because they are unsaturated.

- Organic compounds are **named** according to their functional group and its position in the molecule, the number of carbon atoms in the molecule and the position of any double and triple bonds. The IUPAC rules for nomenclature are used in the naming of organic molecules.
- Many of the **properties** of the hydrocarbons are determined by their **molecular structure**, the **bonds** between atoms and molecules, and their **surface area**.
- The **melting point** and **boiling point** of the hydrocarbons increases as their number of carbon atoms increases.
- The **molecular mass** of the hydrocarbons determines whether they will be in the gaseous, liquid or solid phase at certain temperatures.
- An **alcohol** is an organic compound that contains a **hydroxyl group** (-OH).
- The alcohols have a number of different uses including their use as a solvent, for medicinal purposes and in alcoholic drinks.
- The alcohols share a number of **properties** because of the hydroxyl group. The hydroxyl group affects the **solubility** of the alcohols. Those with shorter carbon chains are generally more soluble, and those with longer chains are less soluble. The strong hydrogen bond between the hydrogen and oxygen atoms in the hydroxyl group gives alcohols a higher melting point and boiling point than other organic compounds. The hydroxyl group also gives the alcohols both acidic and basic properties.
- The **carboxylic acids** are organic acids that contain a **carboxyl group** with the formula -COOH. In a carboxyl group, an oxygen atom is double-bonded to a carbon atom, which is also bonded to a hydroxyl group.
- The carboxylic acids have weak **acidic properties** because the hydrogen atom is able to dissociate from the carboxyl group.
- An **ester** is formed when an alcohol reacts with a carboxylic acid.
- The **amines** are organic compounds that contain an **amino** functional group, which has the formula -NH₂. Some amines belong to the **amino acid** group, which are the building blocks of proteins.
- The **ketones** are a group of compounds that contain a **carbonyl group**, which consists of an oxygen atom that is double-bonded to a carbon atom. In a ketone, the carbonyl group is on the end of the carbon chain.

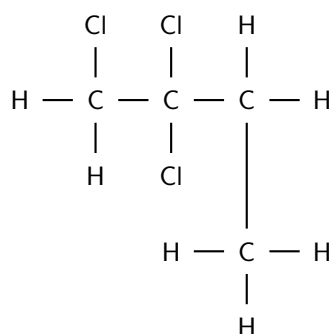


Exercise: Summary exercise

1. Give **one word** for each of the following descriptions:
 - (a) The group of hydrocarbons to which 2-methyl-propene belongs.
 - (b) The name of the functional group that gives alcohols their properties.
 - (c) The group of organic compounds that have acidic properties.
 - (d) The name of the organic compound that is found in vinegar.
 - (e) The name of the organic compound that is found in alcoholic beverages.
2. In each of the following questions, choose the **one correct answer** from the list provided.
 - (a) When 1-propanol is oxidised by acidified potassium permanganate, the possible product formed is...
 - i. propane
 - ii. propanoic acid
 - iii. methyl propanol
 - iv. propyl methanoate

(IEB 2004)

- (b) What is the IUPAC name for the compound represented by the following structural formula?



- i. 1,2,2-trichlorobutane
- ii. 1-chloro-2,2-dichlorobutane
- iii. 1,2,2-trichloro-3-methylpropane
- iv. 1-chloro-2,2-dichloro-3-methylpropane

(IEB 2003)

3. Write balanced equations for the following reactions:

- (a) Ethene reacts with bromine
- (b) Ethyne gas burns in an excess of oxygen
- (c) Ethanoic acid ionises in water

4. The table below gives the boiling point of ten organic compounds.

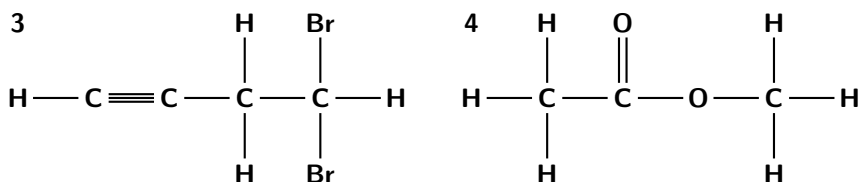
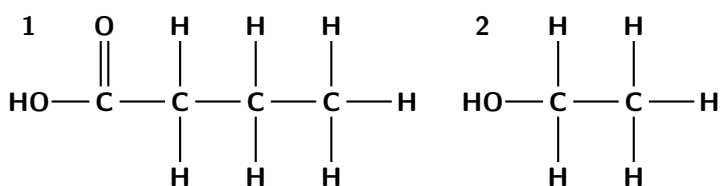
	Compound	Formula	Boiling Point ($^{\circ}\text{C}$)
1	methane	CH_4	-164
2	ethane	C_2H_6	-88
3	propane	C_3H_8	-42
4	butane	C_4H_{10}	0
5	pentane	C_5H_{12}	36
6	methanol	CH_3OH	65
7	ethanol	$\text{C}_2\text{H}_5\text{OH}$	78
8	propan-1-ol	$\text{C}_3\text{H}_7\text{OH}$	98
9	propan-1,2-diol	$\text{CH}_3\text{CHOHCH}_2\text{OH}$	189
10	propan-1,2,3-triol	$\text{CH}_2\text{OHCHOHCH}_2\text{OH}$	290

The following questions refer to the compounds shown in the above table.

- (a) To which homologous series do the following compounds belong?
 - i. Compounds 1,2 and 3
 - ii. Compounds 6,7 and 8
- (b) Which of the above compounds are gases at room temperature?
- (c) What causes the trend of increasing boiling points of compounds 1 to 5?
- (d) Despite the fact that the length of the carbon chain in compounds 8,9 and 10 is the same, the boiling point of propan-1,2,3-triol is much higher than the boiling point of propan-1-ol. What is responsible for this large difference in boiling point?
- (e) Give the IUPAC name and the structural formula of an isomer of butane.
- (f) Which **one** of the above substances is used as a reactant in the preparation of the ester ethylmethanoate?
- (g) Using structural formulae, write an equation for the reaction which produces ethylmethanoate.

(IEB 2004)

5. Refer to the numbered diagrams below and then answer the questions that follow.

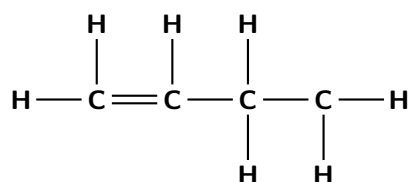


- (a) Which one of the above compounds is produced from the fermentation of starches and sugars in plant matter?
- compound 1
 - compound 2
 - compound 3
 - compound 4
- (b) To which one of the following homologous series does compound 1 belong?
- esters
 - alcohols
 - aldehydes
 - carboxylic acids
- (c) The correct IUPAC name for compound 3 is...
- 1,1-dibromo-3-butyne
 - 4,4-dibromo-1-butyne
 - 2,4-dibromo-1-butyne
 - 4,4-dibromo-1-propyne
- (d) What is the correct IUPAC name for compound 4?
- propanoic acid
 - ethylmethanoate
 - methylethanoate
 - methylpropanoate

IEB 2005

6. Answer the following questions:

- (a) What is a homologous series?
- (b) A mixture of ethanoic acid and methanol is warmed in the presence of concentrated sulphuric acid.
- Using structural formulae, give an equation for the reaction which takes place.
 - What is the IUPAC name of the organic compound formed in this reaction?
- (c) Consider the following unsaturated hydrocarbon:



- i. Give the IUPAC name for this compound.
- ii. Give the balanced equation for the combustion of this compound in excess oxygen.

(IEB Paper 2, 2003)

7. Consider the organic compounds labelled A to E.

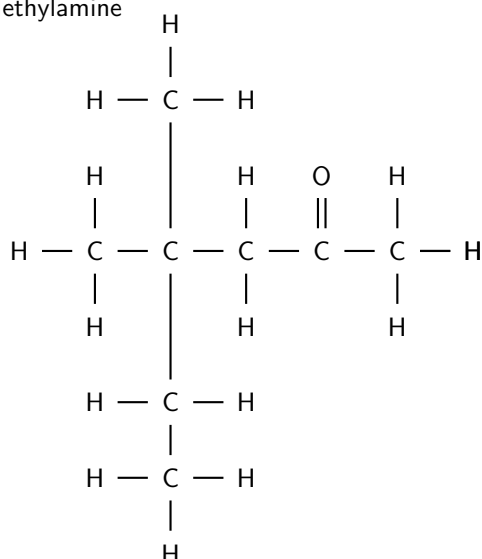
A. $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$

B. C_6H_6

C. $\text{CH}_3\text{-Cl}$

D. Methylamine

E.



- (a) Write a balanced chemical equation for the preparation of compound C using an alkane as one of the reactants.
- (b) Write down the IUPAC name for compound E.
- (c) Write down the structural formula of an isomer of compound A that has only FOUR carbon atoms in the longest chain.
- (d) Write down the structural formula for compound B.

Chapter 2

Organic Macromolecules - Grade 12

As its name suggests, a macromolecule is a large molecule that forms when lots of smaller molecules are joined together. In this chapter, we will be taking a closer look at the structure and properties of different macromolecules, and at how they form.

2.1 Polymers

Some macromolecules are made up of lots of repeating structural units called **monomers**. To put it more simply, a monomer is like a building block. When lots of similar monomers are joined together by covalent bonds, they form a **polymer**. In an **organic polymer**, the monomers are joined by the *carbon* atoms of the polymer 'backbone'. A polymer can also be **inorganic**, in which case there may be atoms such as *silicon* in the place of carbon atoms. The key feature that makes a polymer different from other macromolecules, is the repetition of identical or similar monomers in the polymer chain. The examples shown below will help to make these concepts clearer.



Definition: Polymer

Polymer is a term used to describe large molecules consisting of repeating structural units, or monomers, connected by covalent chemical bonds.

1. Polyethene

Chapter 1 looked at the structure of a group of hydrocarbons called the *alkenes*. One example is the molecule **ethene**. The structural formula of ethene is shown in figure 2.1. When lots of ethene molecules bond together, a polymer called **polyethene** is formed. Ethene is the *monomer* which, when joined to other ethene molecules, forms the *polymer polyethene*. Polyethene is a cheap plastic that is used to make plastic bags and bottles.

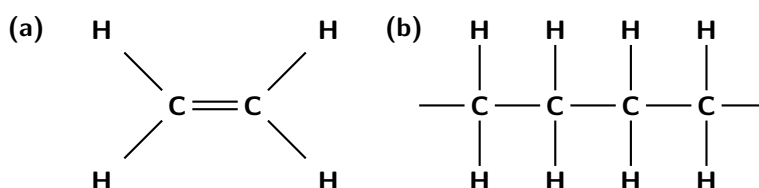


Figure 2.1: (a) Ethene monomer and (b) polyethene polymer

A polymer may be a chain of thousands of monomers, and so it is impossible to draw the entire polymer. Rather, the structure of a polymer can be condensed and represented as

shown in figure 2.2. The monomer is enclosed in brackets and the 'n' represents the number of ethene molecules in the polymer, where 'n' is any whole number. What this shows is that the ethene monomer is repeated an indefinite number of times in a molecule of polyethene.

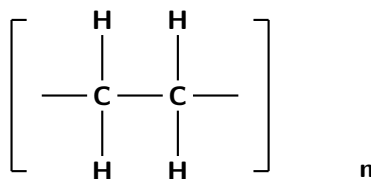


Figure 2.2: A simplified representation of a polyethene molecule

2. Polypropene

Another example of a polymer is *polypropene* (fig 2.3). Polypropene is also a plastic, but is stronger than polyethene and is used to make crates, fibres and ropes. In this polymer, the monomer is the alkene called **propene**.

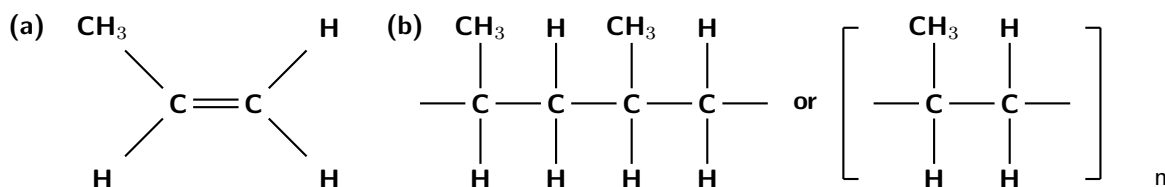


Figure 2.3: (a) Propene monomer and (b) polypropene polymer

2.2 How do polymers form?

Polymers are formed through a process called **polymerisation**, where monomer molecules react together to form a polymer chain. Two types of polymerisation reactions are **addition polymerisation** and **condensation polymerisation**.



Definition: Polymerisation

In chemistry, polymerisation is a process of bonding monomers, or *single units* together through a variety of reaction mechanisms to form longer chains called polymers.

2.2.1 Addition polymerisation

In this type of reaction, monomer molecules are added to a growing polymer chain one at a time. No small molecules are eliminated in the process. An example of this type of reaction is the formation of *polyethene* from *ethene* (fig 2.1). When molecules of ethene are joined to each other, the only thing that changes is that the double bond between the carbon atoms in each ethene monomer is replaced by a single bond so that a new carbon-carbon bond can be formed with the next monomer in the chain. In other words, the monomer is an *unsaturated* compound which, after an addition reaction, becomes a *saturated* compound.



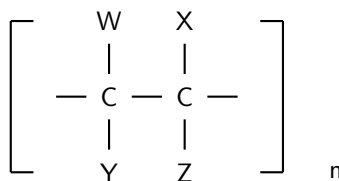
Extension: Initiation, propagation and termination

There are three stages in the process of addition polymerisation. **Initiation** refers to a chemical reaction that triggers off another reaction. In other words, initiation is the starting point of the polymerisation reaction. **Chain propagation** is the part where monomers are continually added to form a longer and longer polymer chain. During chain propagation, it is the reactive end groups of the polymer chain that react in each propagation step, to add a new monomer to the chain. Once a monomer has been added, the reactive part of the polymer is now in this last monomer unit so that propagation will continue. **Termination** refers to a chemical reaction that destroys the reactive part of the polymer chain so that propagation stops.



Worked Example 11: Polymerisation reactions

Question: A polymerisation reaction takes place and the following polymer is formed:



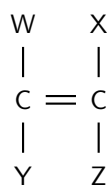
Note: W, X, Y and Z could represent a number of different atoms or combinations of atoms e.g. H, F, Cl or CH₃.

1. Give the structural formula of the monomer of this polymer.
2. To what group of organic compounds does this monomer belong?
3. What type of polymerisation reaction has taken place to join these monomers to form the polymer?

Answer

Step 1 : Look at the structure of the repeating unit in the polymer to determine the monomer.

The monomer is:



Step 2 : Look at the atoms and bonds in the monomer to determine which group of organic compounds it belongs to.

The monomer has a double bond between two carbon atoms. The monomer must be an alkene.

Step 3 : Determine the type of polymerisation reaction.

In this example, unsaturated monomers combine to form a saturated polymer. No atoms are lost or gained for the bonds between monomers to form. They are simply added to each other. This is an addition reaction.

2.2.2 Condensation polymerisation

In this type of reaction, two monomer molecules form a covalent bond and a small molecule such as water is lost in the bonding process. Nearly all biological reactions are of this type. **Polyester** and **nylon** are examples of polymers that form through condensation polymerisation.

1. Polyester

Polyesters are a group of polymers that contain the **ester** functional group in their main chain. Although there are many forms of polyesters, the term *polyester* usually refers to polyethylene terephthalate (PET). PET is made from ethylene glycol (an alcohol) and terephthalic acid (a carboxylic acid). In the reaction, a hydrogen atom is lost from the alcohol, and a hydroxyl group is lost from the carboxylic acid. Together these form one water molecule which is lost during condensation reactions. A new bond is formed between an oxygen and a carbon atom. This bond is called an **ester linkage**. The reaction is shown in figure 2.4.

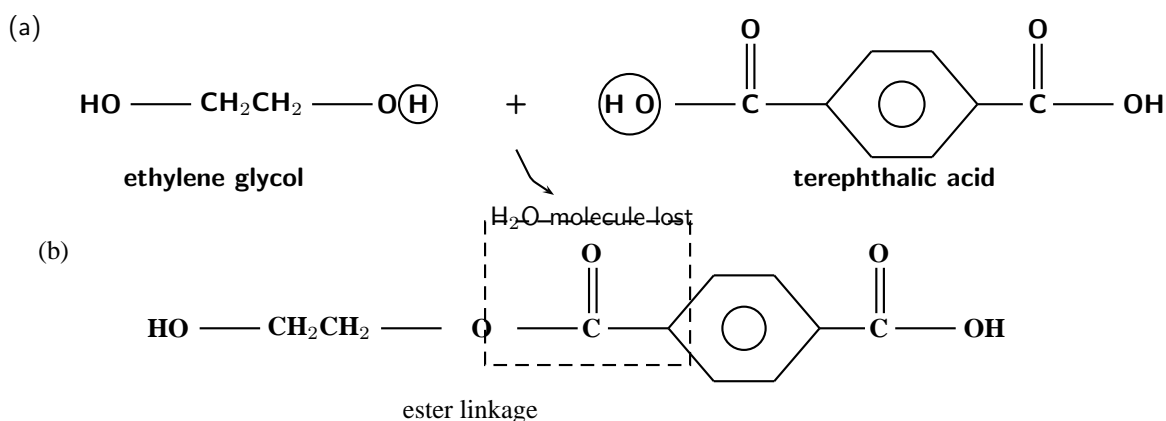
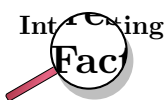


Figure 2.4: An acid and an alcohol monomer react (a) to form a molecule of the polyester 'polyethylene terephthalate' (b).

Polyesters have a number of characteristics which make them very useful. They are resistant to stretching and shrinking, they are easily washed and dry quickly, and they are resistant to mildew. It is for these reasons that polyesters are being used more and more in **textiles**. Polyesters are stretched out into fibres and can then be made into fabric and articles of clothing. In the home, polyesters are used to make clothing, carpets, curtains, sheets, pillows and upholstery.



Polyester is not just a textile. Polyethylene terephthalate is in fact a plastic which can also be used to make plastic drink bottles. Many drink bottles are recycled by being reheated and turned into polyester fibres. This type of recycling helps to reduce disposal problems.

2. Nylon

Nylon was the first polymer to be commercially successful. Nylon replaced silk, and was used to make parachutes during World War 2. Nylon is very strong and resistant, and is used in fishing line, shoes, toothbrush bristles, guitar strings and machine parts to name just a few. Nylon is formed from the reaction of an amine (1,6-diaminohexane) and an

acid monomer (adipic acid) (figure 2.5). The bond that forms between the two monomers is called an **amide linkage**. An amide linkage forms between a nitrogen atom in the amine monomer and the carbonyl group in the carboxylic acid.

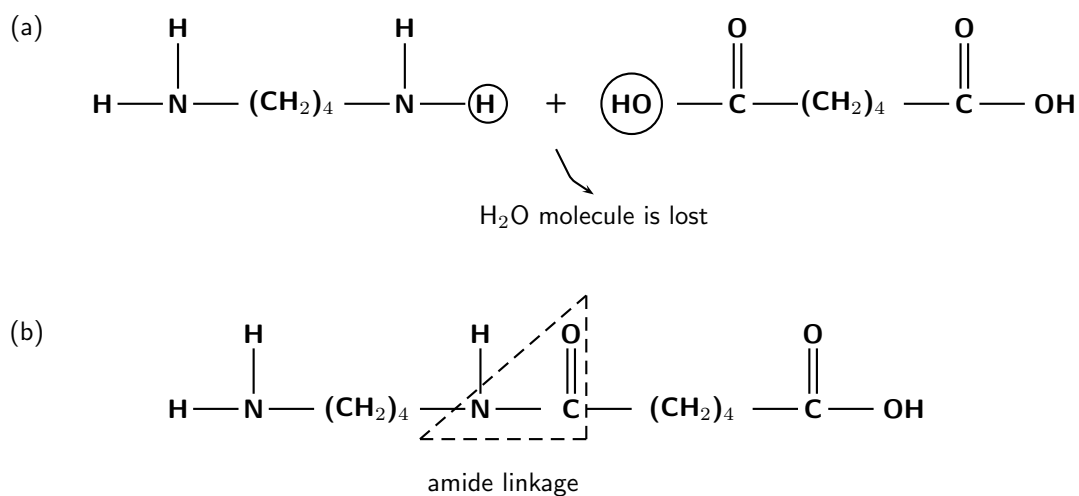
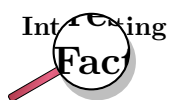


Figure 2.5: An amine and an acid monomer (a) combine to form a section of a nylon polymer (b).

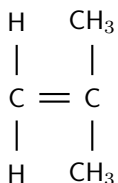


Nylon was first introduced around 1939 and was in high demand to make stockings. However, as World War 2 progressed, nylon was used more and more to make parachutes, and so stockings became more difficult to buy. After the war, when manufacturers were able to shift their focus from parachutes back to stockings, a number of riots took place as women queued to get stockings. In one of the worst disturbances, 40 000 women queued up for 13 000 pairs of stockings, which led to fights breaking out!

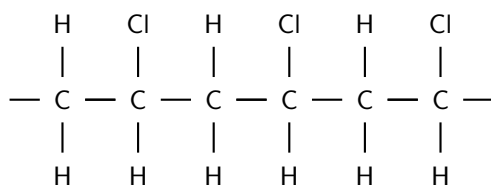


Exercise: Polymers

- The following monomer is a reactant in a polymerisation reaction:



- What is the IUPAC name of this monomer?
 - Give the structural formula of the polymer that is formed in this polymerisation reaction.
 - Is the reaction an addition or condensation reaction?
- The polymer below is the product of a polymerisation reaction.



- (a) Give the structural formula of the monomer in this polymer.
 - (b) What is the name of the monomer?
 - (c) Draw the abbreviated structural formula for the polymer.
 - (d) Has this polymer been formed through an addition or condensation polymerisation reaction?
3. A condensation reaction takes place between methanol and methanoic acid.
- (a) Give the structural formula for...
 - i. methanol
 - ii. methanoic acid
 - iii. the product of the reaction
 - (b) What is the name of the product? (Hint: The product is an ester)

2.3 The chemical properties of polymers

The attractive forces between polymer chains play a large part in determining a polymer's properties. Because polymer chains are so long, these interchain forces are very important. It is usually the side groups on the polymer that determine what types of intermolecular forces will exist. The greater the strength of the intermolecular forces, the greater will be the tensile strength and melting point of the polymer. Below are some examples:

- **Hydrogen bonds between adjacent chains**

Polymers that contain amide or carbonyl groups can form *hydrogen bonds* between adjacent chains. The positive hydrogen atoms in the N-H groups of one chain are strongly attracted to the oxygen atoms (more precisely, the lone-pairs on the oxygen) in the C=O groups on another. Polymers that contain *urea* linkages would fall into this category. The structural formula for urea is shown in figure 2.6. Polymers that contain urea linkages have high tensile strength and a high melting point.

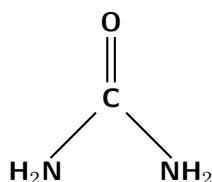


Figure 2.6: The structural formula for urea

- **Dipole-dipole bonds between adjacent chains**

Polyesters have *dipole-dipole bonding* between their polymer chains. Dipole bonding is not as strong as hydrogen bonding, so a polyester's melting point and strength are lower than those of the polymers where there are hydrogen bonds between the chains. However, the weaker bonds between the chains means that polyesters have greater flexibility. The greater the flexibility of a polymer, the more likely it is to be moulded or stretched into fibres.

- **Weak van der Waal's forces**

Other molecules such as ethene do not have a permanent dipole and so the attractive forces between polyethene chains arise from weak *van der Waals* forces. Polyethene therefore has a lower melting point than many other polymers.

2.4 Types of polymers

There are many different types of polymers. Some are organic, while others are inorganic. Organic polymers can be broadly grouped into either synthetic/semi-synthetic (artificial) or biological (natural) polymers. We are going to take a look at two groups of organic polymers: *plastics*, which are usually synthetic or semi-synthetic and *biological macromolecules* which are natural polymers. Both of these groups of polymers play a very important role in our lives.

2.5 Plastics

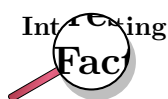
In today's world, we can hardly imagine life without plastic. From cellphones to food packaging, fishing line to plumbing pipes, compact discs to electronic equipment, plastics have become a very important part of our daily lives. "Plastics" cover a range of synthetic and semi-synthetic organic polymers. Their name comes from the fact that they are 'malleable', in other words their shape can be changed and moulded.

**Definition: Plastic**

Plastic covers a range of synthetic or semisynthetic organic polymers. Plastics may contain other substances to improve their performance. Their name comes from the fact that many of them are malleable, in other words they have the property of plasticity.

It was only in the nineteenth century that it was discovered that plastics could be made by chemically changing natural polymers. For centuries before this, only natural organic polymers had been used. Examples of natural organic polymers include *waxes* from plants, *cellulose* (a plant polymer used in fibres and ropes) and *natural rubber* from rubber trees. But in many cases, these natural organic polymers didn't have the characteristics that were needed for them to be used in specific ways. Natural rubber for example, is sensitive to temperature and becomes sticky and smelly in hot weather and brittle in cold weather.

In 1834 two inventors, Friedrich Ludersdorf of Germany and Nathaniel Hayward of the US, independently discovered that adding sulfur to raw rubber helped to stop the material from becoming sticky. After this, Charles Goodyear discovered that heating this modified rubber made it more resistant to abrasion, more elastic and much less sensitive to temperature. What these inventors had done was to improve the properties of a natural polymer so that it could be used in new ways. An important use of rubber now is in vehicle tyres, where these properties of rubber are critically important.



The first true plastic (i.e. one that was not based on any material found in nature) was *Bakelite*, a cheap, strong and durable plastic. Some of these plastics are still used for example in electronic circuit boards, where their properties of insulation and heat resistance are very important.

2.5.1 The uses of plastics

There is such a variety of different plastics available, each having their own specific properties and uses. The following are just a few examples.

- **Polystyrene**

Polystyrene (figure 2.7) is a common plastic that is used in model kits, disposable eating utensils and a variety of other products. In the polystyrene polymer, the monomer is *styrene*, a liquid hydrocarbon that is manufactured from petroleum.

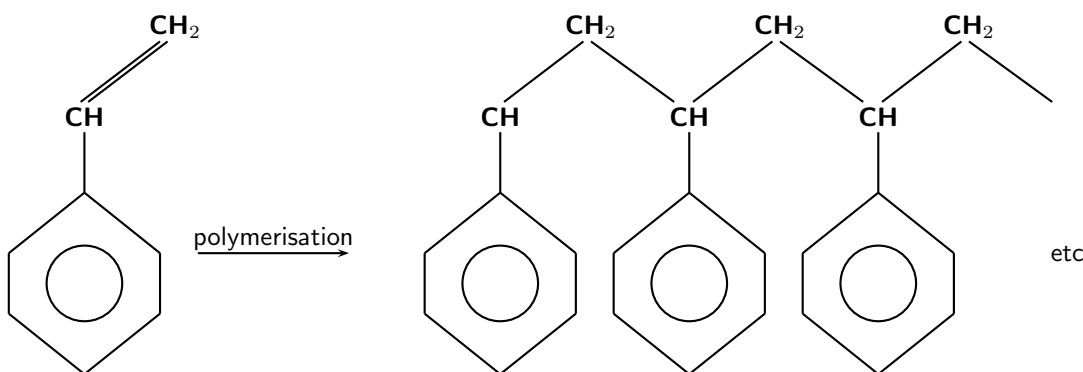


Figure 2.7: The polymerisation of a styrene monomer to form a polystyrene polymer

- **Polyvinylchloride (PVC)**

Polyvinyl chloride (PVC) (figure 2.8) is used in plumbing, gutters, electronic equipment, wires and food packaging. The side chains of PVC contain chlorine atoms, which give it its particular characteristics.

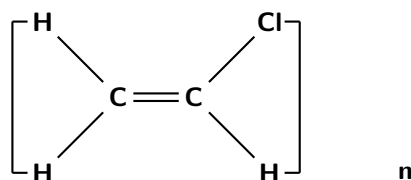
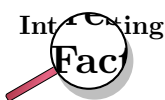


Figure 2.8: Polyvinyl chloride



Many vinyl products have other chemicals added to them to give them particular properties. Some of these chemicals, called additives, can leach out of the vinyl products. In PVC, *plasticizers* are used to make PVC more flexible. Because many baby toys are made from PVC, there is concern that some of these products may leach into the mouths of the babies that are chewing on them. In the USA, most companies have stopped making PVC toys. There are also concerns that some of the plasticizers added to PVC may cause a number of health conditions including cancer.

- **Synthetic rubber**

Another plastic that was critical to the World War 2 effort was *synthetic rubber*, which was produced in a variety of forms. Not only were worldwide natural rubber supplies limited,

but most rubber-producing areas were under Japanese control. Rubber was needed for tyres and parts of war machinery. After the war, synthetic rubber also played an important part in the space race and nuclear arms race.

- **Polyethene/polyethylene (PE)**

Polyethylene (figure 2.1) was discovered in 1933. It is a cheap, flexible and durable plastic and is used to make films and packaging materials, containers and car fittings. One of the most well known polyethylene products is 'Tupperware', the sealable food containers designed by Earl Tupper and promoted through a network of housewives!

- **Polytetrafluoroethylene (PTFE)**

Polytetrafluoroethylene (figure 2.9) is more commonly known as 'Teflon' and is most well known for its use in non-stick frying pans. Teflon is also used to make the breathable fabric Gore-Tex.

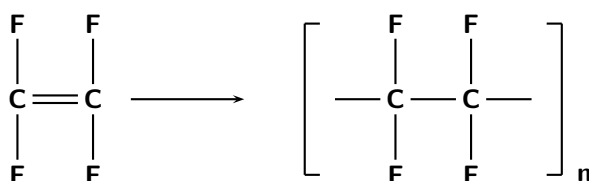


Figure 2.9: A tetra fluoroethylene monomer and polytetrafluoroethylene polymer

Table 2.1 summarises the formulae, properties and uses of some of the most common plastics.

Table 2.1: A summary of the formulae, properties and uses of some common plastics

Name	Formula	Monomer	Properties	Uses
Polyethene (low density)	$-(\text{CH}_2-\text{CH}_2)_n-$	$\text{CH}_2=\text{CH}_2$	soft, waxy solid	film wrap and plastic bags
Polyethene (high density)	$-(\text{CH}_2-\text{CH}_2)_n-$	$\text{CH}_2=\text{CH}_2$	rigid	electrical insulation, bottles and toys
Polypropene	$-(\text{CH}_2-\text{CH}(\text{CH}_3))_n-$	$\text{CH}_2=\text{CHCH}_3$	different grades: some are soft and others hard	carpets and upholstery
Polyvinylchloride (PVC)	$-(\text{CH}_2-\text{CHCl})_n-$	$\text{CH}_2=\text{CHCl}$	strong, rigid	pipes, flooring
Polystyrene	$-(\text{CH}_2-\text{CH}(\text{C}_6\text{H}_5))_n-$	$\text{CH}_2=\text{CHC}_6\text{H}_5$	hard, rigid	toys, packaging
Polytetrafluoroethylene	$-(\text{CF}_2-\text{CF}_2)_n-$	$\text{CF}_2=\text{CF}_2$	resistant, smooth, solid	non-stick surfaces, electrical insulation



Exercise: Plastics

1. It is possible for macromolecules to be composed of more than one type of repeating monomer. The resulting polymer is called a **copolymer**. Varying the monomers that combine to form a polymer, is one way of controlling the properties of the resulting material. Refer to the table below which shows a number of different copolymers of rubber, and answer the questions that follow:

Monomer A	Monomer B	Copolymer	Uses
$\text{H}_2\text{C}=\text{CHCl}$	$\text{H}_2\text{C}=\text{CCl}_2$	Saran	films and fibres
$\text{H}_2\text{C}=\text{CHC}_6\text{H}_5$	$\text{H}_2\text{C}=\text{C}-\text{CH}=\text{CH}_2$	SBR (styrene butadiene rubber)	tyres
$\text{H}_2\text{C}=\text{CHCN}$	$\text{H}_2\text{C}=\text{C}-\text{CH}=\text{CH}_2$	Nitrile rubber	adhesives and hoses
$\text{H}_2\text{C}=\text{C}(\text{CH}_3)_2$	$\text{H}_2\text{C}=\text{C}-\text{CH}=\text{CH}_2$	Butyl rubber	inner tubes
$\text{F}_2\text{C}=\text{CF}(\text{CF}_3)$	$\text{H}_2\text{C}=\text{CHF}$	Viton	gaskets

- Give the structural formula for each of the monomers of nitrile rubber.
 - Give the structural formula of the copolymer viton.
 - In what ways would you expect the properties of SBR to be different from nitrile rubber?
 - Suggest a reason why the properties of these polymers are different.
2. In your home, find as many examples of different types of plastics that you can. Bring them to school and show them to your group. Together, use your examples to complete the following table:

Object	Type of plastic	Properties	Uses

2.5.2 Thermoplastics and thermosetting plastics

A **thermoplastic** is a plastic that can be melted to a liquid when it is heated and freezes to a brittle, glassy state when it is cooled enough. These properties of thermoplastics are mostly due to the fact that the forces between chains are weak. This also means that these plastics can be easily stretched or moulded into any shape. Examples of thermoplastics include nylon, polystyrene, polyethylene, polypropylene and PVC. Thermoplastics are more easily recyclable than some other plastics.

Thermosetting plastics differ from thermoplastics because once they have been formed, they cannot be remelted or remoulded. Examples include bakelite, vulcanised rubber, melanine (used to make furniture), and many glues. Thermosetting plastics are generally stronger than thermoplastics and are better suited to being used in situations where there are high temperatures. They are not able to be recycled. Thermosetting plastics have strong covalent bonds between chains and this makes them very strong.

Activity :: Case Study : Biodegradable plastics

Read the article below and then answer the questions that follow.

Our whole world seems to be wrapped in plastic. Almost every product we buy, most of the food we eat and many of the liquids we drink come encased in plastic. Plastic packaging provides excellent protection for the product, it is cheap to manufacture and seems to last forever. Lasting forever, however, is proving to be a major environmental problem. Another problem is that traditional plastics are manufactured from non-renewable resources - oil, coal and natural gas. In an effort to overcome these problems, researchers and engineers have been trying to develop biodegradable plastics that are made from renewable resources, such as plants.

The term biodegradable means that a substance can be broken down into simpler substances by the activities of living organisms, and therefore is unlikely to remain in the environment. The reason most plastics are not biodegradable is because their long polymer molecules are too large and too tightly bonded together to be broken apart and used by decomposer organisms. However, plastics based on natural plant polymers that come from wheat or corn starch have molecules that can be more easily broken down by microbes.

Starch is a natural polymer. It is a white, granular carbohydrate produced by plants during photosynthesis and it serves as the plant's energy store. Many plants contain large amounts of starch. Starch can be processed directly into a bioplastic but, because it is soluble in water, articles made from starch will swell and deform when exposed to moisture, and this limits its use. This problem can be overcome by changing starch into a different polymer. First, starch is harvested from corn, wheat or potatoes, then microorganisms transform it into lactic acid, a monomer. Finally, the lactic acid is chemically treated to cause the molecules of lactic acid to link up into long chains or polymers, which bond together to form a plastic called polylactide (PLA).

PLA can be used for products such as plant pots and disposable nappies. It has been commercially available in some countries since 1990, and certain blends have proved successful in medical implants, sutures and drug delivery systems because they are able to dissolve away over time. However, because PLA is much more expensive than normal plastics, it has not become as popular as one would have hoped.

Questions

1. In your own words, explain what is meant by a 'biodegradable plastic'.
 2. Using your knowledge of chemical bonding, explain why some polymers are biodegradable and others are not.
 3. Explain why lactic acid is a more useful monomer than starch, when making a biodegradable plastic.
 4. If you were a consumer (shopper), would you choose to buy a biodegradable plastic rather than another? Explain your answer.
 5. What do you think could be done to make biodegradable plastics more popular with consumers?
-

2.5.3 Plastics and the environment

Although plastics have had a huge impact globally, there is also an environmental price that has to be paid for their use. The following are just some of the ways in which plastics can cause damage to the environment.

1. Waste disposal

Plastics are not easily broken down by micro-organisms and therefore most are not easily biodegradable. This leads to waste disposal problems.

2. Air pollution

When plastics burn, they can produce toxic gases such as carbon monoxide, hydrogen cyanide and hydrogen chloride (particularly from PVC and other plastics that contain chlorine and nitrogen).

3. Recycling

It is very difficult to recycle plastics because each type of plastic has different properties and so different recycling methods may be needed for each plastic. However, attempts are being made to find ways of recycling plastics more effectively. Some plastics can be remelted and re-used, while others can be ground up and used as a filler. However,

one of the problems with recycling plastics is that they have to be sorted according to plastic *type*. This process is difficult to do with machinery, and therefore needs a lot of labour. Alternatively, plastics should be re-used. In many countries, including South Africa, shoppers must now pay for plastic bags. This encourages people to collect and re-use the bags they already have.

Activity :: Case Study : Plastic pollution in South Africa

Read the following extract, taken from 'Planet Ark' (September 2003), and then answer the questions that follow.

South Africa launches a programme this week to exterminate its "national flower" - the millions of used plastic bags that litter the landscape.

Beginning on Friday, plastic shopping bags used in the country must be both thicker and more recyclable, a move officials hope will stop people from simply tossing them away. "Government has targeted plastic bags because they are the most visible kind of waste," said Phindile Makwakwa, spokeswoman for the Department of Environmental Affairs and Tourism. "But this is mostly about changing people's mindsets about the environment."

South Africa is awash in plastic pollution. Plastic bags are such a common eyesore that they are dubbed "roadside daisies" and referred to as the national flower. Bill Naude of the Plastics Federation of South Africa said the country used about eight billion plastic bags annually, a figure which could drop by 50 percent if the new law works.

It is difficult sometimes to imagine exactly how much waste is produced in our country every year. Where does all of this go to? You are going to do some simple calculations to try to estimate the volume of plastic packets that is produced in South Africa every year.

1. Take a plastic shopping packet and squash it into a tight ball.
 - (a) Measure the approximate length, breadth and depth of your squashed plastic bag.
 - (b) Calculate the approximate volume that is occupied by the packet.
 - (c) Now calculate the approximate volume of your classroom by measuring its length, breadth and height.
 - (d) Calculate the number of squashed plastic packets that would fit into a classroom of this volume.
 - (e) If South Africa produces an average of 8 billion plastic bags each year, how many classrooms would be filled if all of these bags were thrown away and not re-used?
 2. What has South Africa done to try to reduce the number of plastic bags that are produced?
 3. Do you think this has helped the situation?
 4. What can *you* do to reduce the amount of plastic that you throw away?
-

2.6 Biological Macromolecules

A *biological macromolecule* is one that is found in living organisms. Biological macromolecules include molecules such as carbohydrates, proteins and nucleic acids. Lipids are also biological macromolecules. They are essential for all known forms of life to survive.

**Definition: Biological macromolecule**

A biological macromolecule is a polymer that occurs naturally in living organisms. These molecules are essential to the survival of life.

2.6.1 Carbohydrates

Carbohydrates include the sugars and their polymers. One key characteristic of the carbohydrates is that they contain only the elements carbon, hydrogen and oxygen. In the carbohydrate monomers, every carbon except one has a hydroxyl group attached to it, and the remaining carbon atom is double bonded to an oxygen atom to form a carbonyl group. One of the most important monomers in the carbohydrates is **glucose** (figure 2.10). The glucose molecule can exist in an open-chain (acyclic) and ring (cyclic) form.

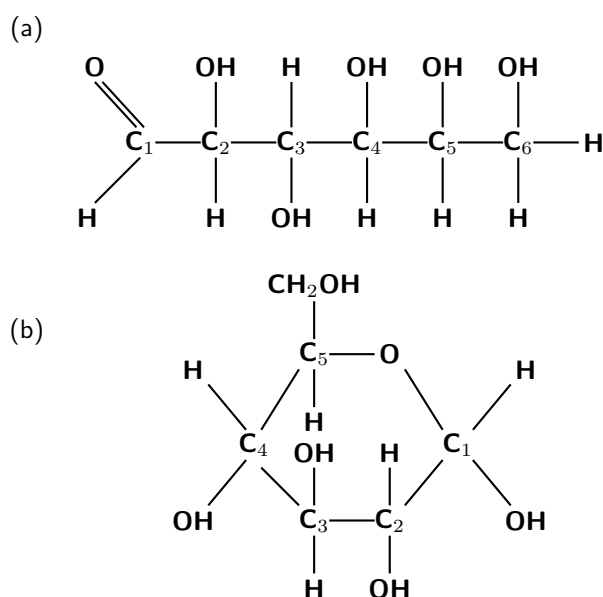


Figure 2.10: The open chain (a) and cyclic (b) structure of a glucose molecule

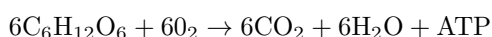
Glucose is produced during **photosynthesis**, which takes place in plants. During photosynthesis, sunlight (solar energy), water and carbon dioxide are involved in a chemical reaction that produces glucose and oxygen. This glucose is stored in various ways in the plant.

The photosynthesis reaction is as follows:



Glucose is an important source of **energy** for both the plant itself, and also for the other animals and organisms that may feed on it. Glucose plays a critical role in **cellular respiration**, which is a chemical reaction that occurs in the cells of all living organisms. During this reaction, glucose and oxygen react to produce carbon dioxide, water and Adenosine Triphosphate (ATP). ATP is a molecule that cells use for energy so that the body's cells can function normally. The purpose of *eating* then, is to obtain glucose which the body can then convert into the ATP it needs to be able to survive.

The reaction for cellular respiration is as follows:



We don't often eat glucose in its simple form. More often, we eat complex carbohydrates that our bodies have to break down into individual glucose molecules before they can be used in cellular respiration. These complex carbohydrates are polymers, which form through condensation polymerisation reactions (figure 2.11). *Starch* and *cellulose* are two examples of carbohydrates that are polymers composed of glucose monomers.

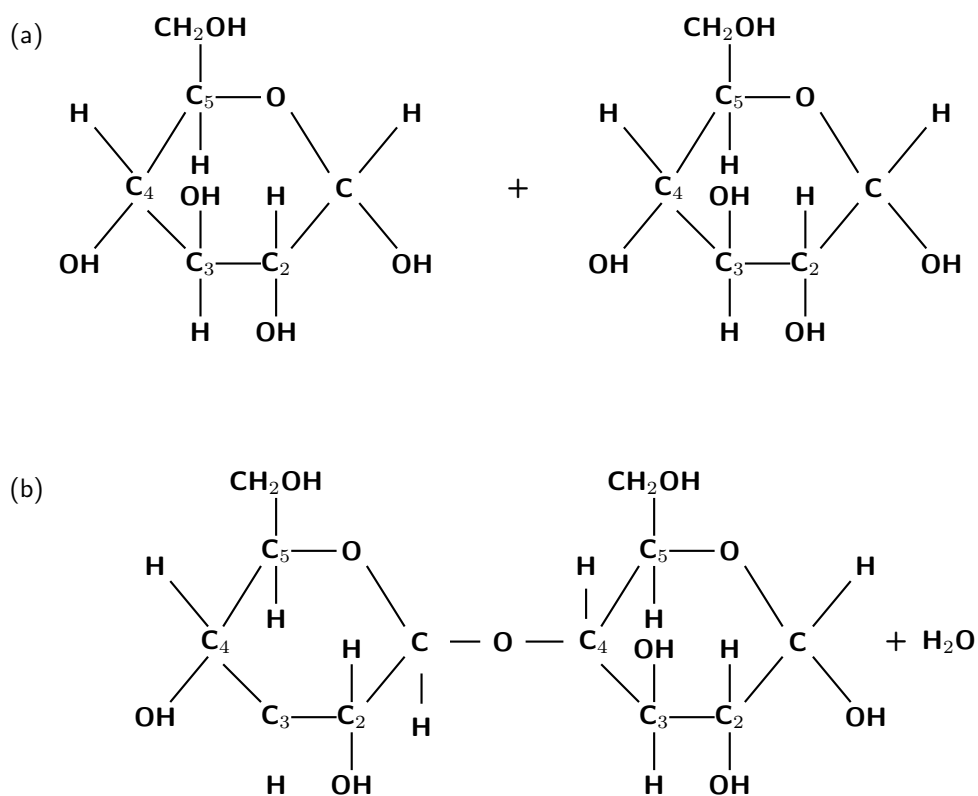


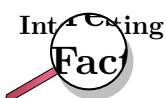
Figure 2.11: Two glucose monomers (a) undergo a condensation reaction to produce a section of a carbohydrate polymer (b). One molecule of water is produced for every two monomers that react.

- **Starch**

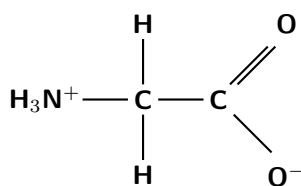
Starch is used by plants to store excess glucose, and consists of long chains of glucose monomers. Potatoes are made up almost entirely of starch. This is why potatoes are such a good source of energy. Animals are also able to store glucose, but in this case it is stored as a compound called **glycogen**, rather than as starch.

- **Cellulose**

Cellulose is also made up of chains of glucose molecules, but the bonding between the polymers is slightly different from that in starch. Cellulose is found in the cell walls of plants and is used by plants as a building material.



It is very difficult for animals to digest the cellulose in plants that they may have been feeding on. However, fungi and some protozoa are able to break down cellulose. Many animals, including termites and cows, use these organisms to break cellulose down into glucose, which they can then use more easily.



glycine

When two amino acid monomers are close together, they may be joined to each other by **peptide bonds** (figure 2.13) to form a **polypeptide** chain. . The reaction is a condensation reaction. Polypeptides can vary in length from a few amino acids to a thousand or more. The polypeptide chains are then joined to each other in different ways to form a **protein**. It is the sequence of the amino acids in the polymer that gives a protein its particular properties.

The sequence of the amino acids in the chain is known as the protein's **primary structure**. As the chain grows in size, it begins to twist, curl and fold upon itself. The different parts of the polypeptide are held together by hydrogen bonds, which form between hydrogen atoms in one part of the chain and oxygen or nitrogen atoms in another part of the chain. This is known as the **secondary structure** of the protein. Sometimes, in this coiled helical structure, bonds may form between the side chains (R groups) of the amino acids. This results in even more irregular contortions of the protein. This is called the **tertiary structure** of the protein.

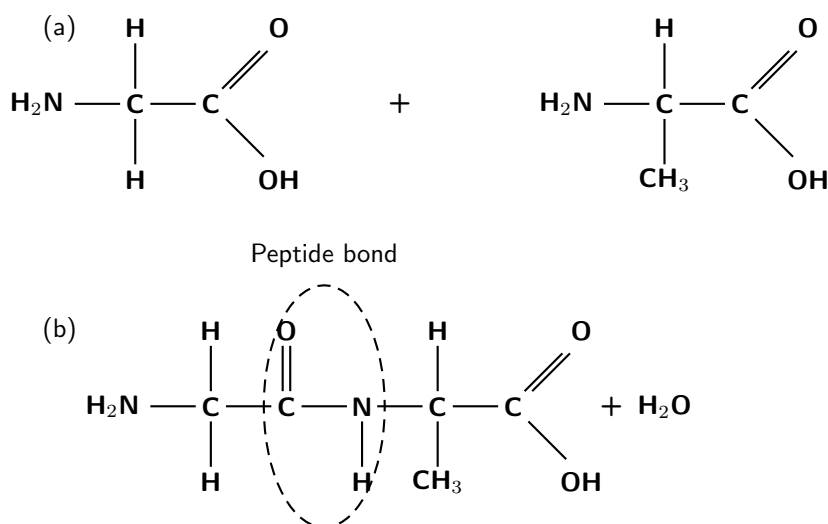
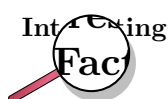


Figure 2.13: Two amino acids (glycine and alanine) combine to form part of a polypeptide chain. The amino acids are joined by a peptide bond between a carbon atom of one amino acid and a nitrogen atom of the other amino acid.



There are twenty different amino acids that exist in nature. All cells, both plant and animal, build their proteins from only twenty amino acids. At first, this seems like a very small number, especially considering the huge number

of different proteins that exist. However, if you consider that most proteins are made up of polypeptide chains that contain at least 100 amino acids, you will start to realise the endless possible combinations of amino acids that are available.

The functions of proteins

Proteins have a number of functions in living organisms.

- *Structural proteins* such as collagen in animal connective tissue and keratin in hair, horns and feather quills, all provide support.
- *Storage proteins* such as albumin in egg white provide a source of energy. Plants store proteins in their seeds to provide energy for the new growing plant.
- *Transport proteins* transport other substances in the body. Haemoglobin in the blood for example, is a protein that contains iron. Haemoglobin has an affinity (attraction) for oxygen and so this is how oxygen is transported around the body in the blood.
- *Hormonal proteins* coordinate the body's activities. Insulin for example, is a hormonal protein that controls the sugar levels in the blood.
- *Enzymes* are chemical catalysts and speed up chemical reactions. Digestive enzymes such as amylase in your saliva, help to break down polymers in food. Enzymes play an important role in all cellular reactions such as respiration, photosynthesis and many others.

Activity :: Research Project : Macromolecules in our daily diet

1. In order to keep our bodies healthy, it is important that we eat a balanced diet with the right amounts of carbohydrates, proteins and fats. Fats are an important source of energy, they provide insulation for the body, and they also provide a protective layer around many vital organs. Our bodies also need certain essential vitamins and minerals. Most food packaging has a label that provides this information.

Choose a number of different food items that you eat. Look at the food label for each, and then complete the following table:

Food	Carbohydrates (%)	Proteins (%)	Fats (%)

- (a) Which food type contains the largest proportion of protein?
 - (b) Which food type contains the largest proportion of carbohydrates?
 - (c) Which of the food types you have listed would you consider to be the 'healthiest'? Give a reason for your answer.
2. In an effort to lose weight, many people choose to *diet*. There are many diets on offer, each of which is based on particular theories about how to lose weight most effectively. Look at the list of diets below:
 - Vegetarian diet
 - Low fat diet

- Atkin's diet
- Weight Watchers

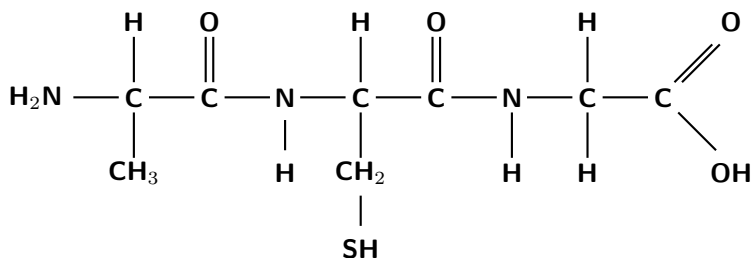
For each of these diets, answer the following questions:

- What theory of weight loss does each type of diet propose?
 - What are the *benefits* of the diet?
 - What are the potential *problems* with the diet?
-
-



Exercise: Carbohydrates and proteins

- Give the structural formula for each of the following:
 - A polymer chain, consisting of three glucose molecules.
 - A polypeptide chain, consisting of two molecules of alanine and one molecule of serine.
- Write balanced equations to show the polymerisation reactions that produce the polymers described above.
- The following polypeptide is the end product of a polymerisation reaction:



- Give the structural formula of the monomers that make up the polypeptide.
 - On the structural formula of the first monomer, label the amino group and the carboxyl group.
 - What is the chemical formula for the carbon side chain in the second monomer?
 - Name the bond that forms between the monomers of the polypeptide.
-

2.6.3 Nucleic Acids

You will remember that we mentioned earlier that each protein is different because of its unique sequence of amino acids. But what controls how the amino acids arrange themselves to form the specific proteins that are needed by an organism? This task is for the **gene**. A gene contains DNA (deoxyribonucleic acid) which is a polymer that belongs to a class of compounds called the **nucleic acids**. DNA is the genetic material that organisms inherit from their parents. It is DNA that provides the genetic coding that is needed to form the specific proteins that an organism needs. Another nucleic acid is RNA (ribonucleic acid). The diagram in figure 2.14 shows an RNA molecule.

The DNA polymer is made up of monomers called **nucleotides**. Each nucleotide has three parts: a sugar, a phosphate and a nitrogenous base. DNA is a double-stranded helix (a helix is basically

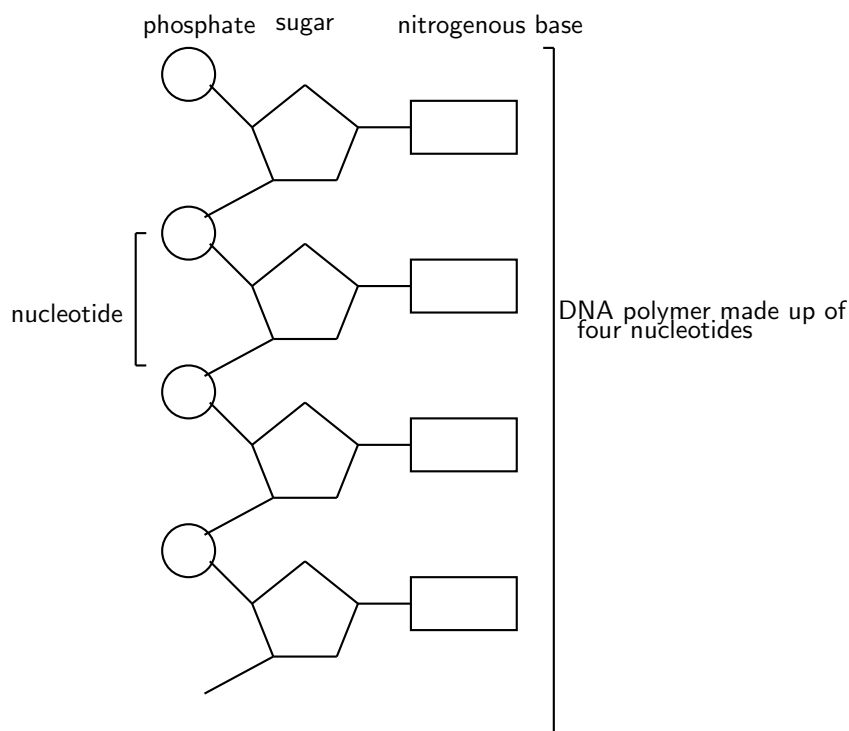
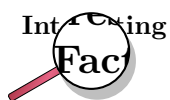


Figure 2.14: Nucleotide monomers make up the RNA polymer

a coil). Or you can think of it as two RNA molecules bonded together.

There are five different nitrogenous bases: adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). It is the sequence of the nitrogenous bases in a DNA polymer that will determine the genetic code for that organism. Three consecutive nitrogenous bases provide the coding for one amino acid. So, for example, if the nitrogenous bases on three nucleotides are *uracil*, *cytosine* and *uracil* (in that order), one **serine** amino acid will become part of the polypeptide chain. The polypeptide chain is built up in this way until it is long enough (and with the right amino acid sequence) to be a protein. Since proteins control much of what happens in living organisms, it is easy to see how important nucleic acids are as the starting point of this process.



A single defect in even one nucleotide, can be devastating to an organism. One example of this is a disease called **sickle cell anaemia**. Because of one wrong nucleotide in the genetic code, the body produces a protein called **sickle haemoglobin**. Haemoglobin is the protein in red blood cells that helps to transport oxygen around the body. When sickle haemoglobin is produced, the red blood cells change shape. This process damages the red blood cell membrane, and can cause the cells to become stuck in blood vessels. This then means that the red blood cells, which are carrying oxygen, can't get to the tissues where they are needed. This can cause serious organ damage. Individuals who have sickle cell anaemia generally have a lower life expectancy.

Table 2.2 shows some other examples of genetic coding for different amino acids.

Table 2.2: Nitrogenous base sequences and the corresponding amino acid

Nitrogenous base sequence	Amino acid
UUU	Phenylalanine
CUU	Leucine
UCU	Serine
UAU	Tyrosine
UGU	Cysteine
GUU	Valine
GCU	Alanine
GGU	Glycine



Exercise: Nucleic acids

- For each of the following, say whether the statement is **true** or **false**. If the statement is *false*, give a reason for your answer.
 - Deoxyribonucleic acid (DNA) is an example of a *polymer* and a nucleotide is an example of a *monomer*.
 - Thymine and uracil are examples of nucleotides.
 - A person's DNA will determine what proteins their body will produce, and therefore what characteristics they will have.
 - An amino acid is a protein monomer.
 - A polypeptide that consists of five amino acids, will also contain five nucleotides.
- For each of the following sequences of nitrogenous bases, write down the amino acid/s that will be part of the polypeptide chain.
 - UUU
 - UCUUUU
 - GGUUAUGUUGGU
- A polypeptide chain consists of three amino acids. The sequence of nitrogenous bases in the nucleotides of the DNA is GCUGGUGCU. Give the structural formula of the polypeptide.

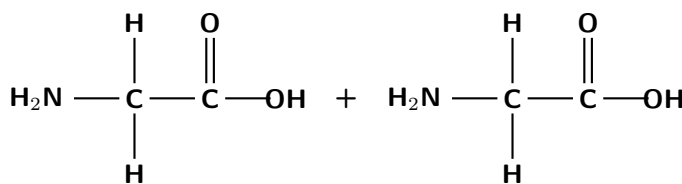
2.7 Summary

- A **polymer** is a macromolecule that is made up of many repeating structural units called **monomers** which are joined by covalent bonds.
- Polymers that contain carbon atoms in the main chain are called **organic polymers**.
- Organic polymers can be divided into **natural organic polymers** (e.g. natural rubber) or **synthetic organic polymers** (e.g. polystyrene).
- The polymer **polyethene** for example, is made up of many ethene monomers that have been joined into a polymer chain.
- Polymers form through a process called **polymerisation**.
- Two examples of polymerisation reactions are **addition** and **condensation** reactions.
- An **addition reaction** occurs when unsaturated monomers (e.g. alkenes) are added to each other one by one. The breaking of a double bond between carbon atoms in the monomer, means that a bond can form with the next monomer. The polymer **polyethene** is formed through an addition reaction.

- In a **condensation reaction**, a molecule of water is released as a product of the reaction. The water molecule is made up of atoms that have been lost from each of the monomers. Polyesters and nylon are polymers that are produced through a condensation reaction.
- The **chemical properties** of polymers (e.g. tensile strength and melting point) are determined by the types of atoms in the polymer, and by the strength of the bonds between adjacent polymer chains. The stronger the bonds, the greater the strength of the polymer, and the higher its melting point.
- One group of synthetic organic polymers, are the **plastics**.
- **Polystyrene** is a plastic that is made up of styrene monomers. Polystyrene is used a lot in packaging.
- **Polyvinyl chloride** (PVC) consists of vinyl chloride monomers. PVC is used to make pipes and flooring.
- **Polyethene**, or **polyethylene**, is made from ethene monomers. Polyethene is used to make film wrapping, plastic bags, electrical insulation and bottles.
- **Polytetrafluoroethylene** is used in non-stick frying pans and electrical insulation.
- A **thermoplastic** can be heated and melted to a liquid. It freezes to a brittle, glassy state when cooled very quickly. Examples of thermoplastics are polyethene and PVC.
- A **thermoset** plastic cannot be melted or re-shaped once formed. Examples of thermoset plastics are vulcanised rubber and melamine.
- It is not easy to **recycle** all plastics, and so they create environmental problems.
- Some of these **environmental problems** include issues of waste disposal, air pollution and recycling.
- A **biological macromolecule** is a polymer that occurs naturally in living organisms.
- Examples of biological macromolecules include **carbohydrates** and **proteins**, both of which are essential for life to survive.
- Carbohydrates include the **sugars** and their polymers, and are an important source of **energy** in living organisms.
- **Glucose** is a carbohydrate monomer. Glucose is the molecule that is needed for **cellular respiration**.
- The glucose monomer is also a building block for carbohydrate polymers such as **starch**, **glycogen** and **cellulose**.
- **Proteins** have a number of important functions. These include their roles in structures, transport, storage, hormonal proteins and enzymes.
- A protein consists of monomers called **amino acids**, which are joined by **peptide bonds**.
- A protein has a **primary**, **secondary** and **tertiary** structure.
- An amino acid is an organic molecule, made up of a **carboxyl** and an **amino** group, as well as a carbon **side chain** of varying lengths.
- It is the **sequence** of amino acids that determines the nature of the protein.
- It is the **DNA** of an organism that determines the order in which amino acids combine to make a protein.
- DNA is a **nucleic acid**. DNA is a polymer, and is made up of monomers called **nucleotides**.
- Each nucleotide consists of a **sugar**, a **phosphate** and a **nitrogenous base**. It is the sequence of the nitrogenous bases that provides the 'code' for the arrangement of the amino acids in a protein.


Exercise: Summary exercise

- Give one word for each of the following descriptions:
 - A chain of monomers joined by covalent bonds.
 - A polymerisation reaction that produces a molecule of water for every two monomers that bond.
 - The bond that forms between an alcohol and a carboxylic acid monomer during a polymerisation reaction.
 - The name given to a protein monomer.
 - A six-carbon sugar monomer.
 - The monomer of DNA, which determines the sequence of amino acids that will make up a protein.
- For each of the following questions, choose the one correct answer from the list provided.
 - A polymer is made up of monomers, each of which has the formula $\text{CH}_2=\text{CHCN}$. The formula of the polymer is:
 - $-(\text{CH}_2=\text{CHCN})_n-$
 - $-(\text{CH}_2-\text{CHCN})_n-$
 - $-(\text{CH}-\text{CHCN})_n-$
 - $-(\text{CH}_3-\text{CHCN})_n-$
 - A polymer has the formula $-\text{[CO}(\text{CH}_2)_4\text{CO-NH}(\text{CH}_2)_6\text{NH}]_n-$. Which of the following statements is **true**?
 - The polymer is the product of an addition reaction.
 - The polymer is a polyester.
 - The polymer contains an amide linkage.
 - The polymer contains an ester linkage.
 - Glucose...
 - is a monomer that is produced during cellular respiration
 - is a sugar polymer
 - is the monomer of starch
 - is a polymer produced during photosynthesis
- The following monomers are involved in a polymerisation reaction:



- Give the structural formula of the polymer that is produced.
 - Is the reaction an addition or condensation reaction?
 - To what group of organic compounds do the two monomers belong?
 - What is the name of the monomers?
 - What type of bond forms between the monomers in the final polymer?
- The table below shows the melting point for three plastics. Suggest a reason why the melting point of PVC is *higher* than the melting point for polyethene, but *lower* than that for polyester.

Plastic	Melting point ($^{\circ}\text{C}$)
Polyethene	105 - 115
PVC	212
Polyester	260

5. An amino acid has the formula $\text{H}_2\text{NCH}(\text{CH}_2\text{CH}_2\text{SCH}_3)\text{COOH}$.
- Give the structural formula of this amino acid.
 - What is the chemical formula of the carbon side chain in this molecule?
 - Are there any peptide bonds in this molecule? Give a reason for your answer.
-

Chapter 3

Reaction Rates - Grade 12

3.1 Introduction

Before we begin this section, it might be useful to think about some different types of reactions and how quickly or slowly they occur.

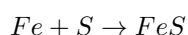


Exercise: Thinking about reaction rates

Think about each of the following reactions:

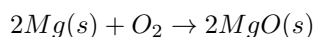
- rusting of metals
 - photosynthesis
 - weathering of rocks (e.g. limestone rocks being weathered by water)
 - combustion
1. For each of the reactions above, write a chemical equation for the reaction that takes place.
 2. How fast is each of these reactions? Rank them in order from the fastest to the slowest.
 3. How did you decide which reaction was the fastest and which was the slowest?
 4. Try to think of some other examples of chemical reactions. How fast or slow is each of these reactions, compared with those listed earlier?
-

In a chemical reaction, the substances that are undergoing the reaction are called the **reactants**, while the substances that form as a result of the reaction are called the **products**. The **reaction rate** describes how quickly or slowly the reaction takes place. So how do we know whether a reaction is slow or fast? One way of knowing is to look either at how quickly the *reactants are used up* during the reaction or at how quickly the *product forms*. For example, iron and sulfur react according to the following equation:



In this reaction, we can observe the speed of the reaction by measuring how long it takes before there is no iron or sulfur left in the reaction vessel. In other words, the reactants have been used up. Alternatively, one could see how quickly the iron sulfide product forms. Since iron sulfide looks very different from either of its reactants, this is easy to do.

In another example:



In this case, the reaction rate depends on the speed at which the reactants (solid magnesium and oxygen gas) are used up, or the speed at which the product (magnesium oxide) is formed.



Definition: Reaction rate

The *rate of a reaction* describes how quickly *reactants are used up* or how quickly *products are formed* during a chemical reaction. The units used are: moles per second (mols/second or $\text{mol}\cdot\text{s}^{-1}$).

The average rate of a reaction is expressed as the number of moles of reactant used up, divided by the total reaction time, or as the number of moles of product formed, divided by the reaction time. Using the magnesium reaction shown earlier:

$$\text{Average reaction rate of Mg} = \frac{\text{moles Mg used}}{\text{reaction time (s)}}$$

or

$$\text{Average reaction rate of O}_2 = \frac{\text{moles O}_2 \text{ used}}{\text{reaction time (s)}}$$

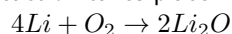
or

$$\text{Average reaction rate of MgO} = \frac{\text{moles MgO produced}}{\text{reaction time (s)}}$$



Worked Example 12: Reaction rates

Question: The following reaction takes place:



After two minutes, 4 g of Lithium has been used up. Calculate the rate of the reaction.

Answer

Step 1 : Calculate the number of moles of Lithium that are used up in the reaction.

$$n = \frac{m}{M} = \frac{4}{6.94} = 0.58 \text{ mols}$$

Step 2 : Calculate the time (in seconds) for the reaction.

$$t = 2 \times 60 = 120 \text{ s}$$

Step 3 : Calculate the rate of the reaction.

Rate of reaction =

$$\frac{\text{moles of Lithium used}}{\text{time}} = \frac{0.58}{120} = 0.005$$

The rate of the reaction is $0.005 \text{ mol}\cdot\text{s}^{-1}$

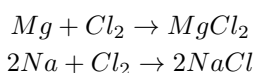


Exercise: Reaction rates

1. A number of different reactions take place. The table below shows the number of moles of reactant that are used up in a particular time for each reaction.

Reaction	Mols used up	Time	Reaction rate
1	2	30 secs	
2	5	2 mins	
3	1	1.5 mins	
4	3.2	1.5 mins	
5	5.9	30 secs	

- (a) Complete the table by calculating the rate of each reaction.
 (b) Which is the *fastest* reaction?
 (c) Which is the *slowest* reaction?
2. Two reactions occur simultaneously in separate reaction vessels. The reactions are as follows:



After 1 minute, 2 g of MgCl_2 have been produced in the first reaction.

- (a) How many moles of MgCl_2 are produced after 1 minute?
 (b) Calculate the rate of the reaction, using the amount of product that is produced.
 (c) Assuming that the second reaction also proceeds at the same rate, calculate...
 i. the number of moles of NaCl produced after 1 minute.
 ii. the mass (in g) of sodium that is needed for this reaction to take place.
-

3.2 Factors affecting reaction rates

Several factors affect the rate of a reaction. It is important to know these factors so that reaction rates can be controlled. This is particularly important when it comes to industrial reactions, so that productivity can be maximised. The following are some of the factors that affect the rate of a reaction.

1. Nature of reactants

Substances have different chemical properties and therefore react differently and at different rates.

2. Concentration (or pressure in the case of gases)

As the concentration of the reactants increases, so does the reaction rate.

3. Temperature

If the temperature of the reaction increases, so does the rate of the reaction.

4. Catalyst

Adding a catalyst increases the reaction rate.

5. Surface area of solid reactants

Increasing the surface area of the reactants (e.g. if a solid reactant is broken or ground up into smaller pieces) will increase the reaction rate.

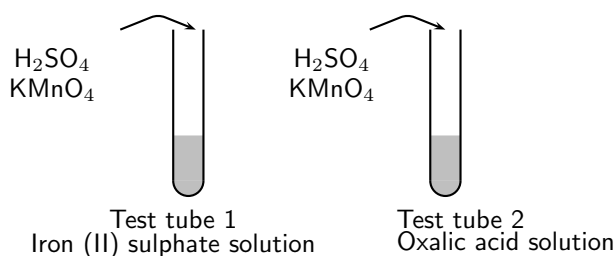
Activity :: Experiment : The nature of reactants.

Aim:

To determine the effect of the nature of reactants on the rate of a reaction.

Apparatus:

Oxalic acid ((COOH)₂), iron(II) sulphate (FeSO₄), potassium permanganate (KMnO₄), concentrated sulfuric acid (H₂SO₄), spatula, test tubes, medicine dropper, glass beaker and glass rod.



Method:

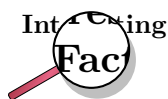
1. In the first test tube, prepare an iron (II) sulphate solution by dissolving about two spatula points of iron (II) sulphate in 10 cm³ of water.
2. In the second test tube, prepare a solution of oxalic acid in the same way.
3. Prepare a solution of sulfuric acid by adding 1 cm³ of the concentrated acid to about 4 cm³ of water. Remember always to add the *acid to the water*, and never the other way around.
4. Add 2 cm³ of the sulfuric acid solution to the iron(II) and oxalic acid solutions respectively.
5. Using the medicine dropper, add a few drops of potassium permanganate to the two test tubes. Once you have done this, observe how quickly each solution discolours the potassium permanganate solution.

Results:

- You should have seen that the oxalic acid solution discolours the potassium permanganate much more slowly than the iron(II) sulphate.
- It is the oxalate ions (C₂O₄²⁻) and the Fe²⁺ ions that cause the discolouration. It is clear that the Fe²⁺ ions act much more quickly than the C₂O₄²⁻ ions. The reason for this is that there are no covalent bonds to be broken in the ions before the reaction can take place. In the case of the oxalate ions, covalent bonds between carbon and oxygen atoms must be broken first.

Conclusions:

The nature of the reactants can affect the rate of a reaction.



Oxalic acids are abundant in many plants. The leaves of the tea plant (*Camellia sinensis*) contain very high concentrations of oxalic acid relative to other plants. Oxalic acid also occurs in small amounts in foods such as parsley, chocolate, nuts and berries. Oxalic acid irritates the lining of the gut when it is eaten, and can be fatal in very large doses.

Activity :: Experiment : Surface area and reaction rates.

Marble (CaCO_3) reacts with hydrochloric acid (HCl) to form calcium chloride, water and carbon dioxide gas according to the following equation:

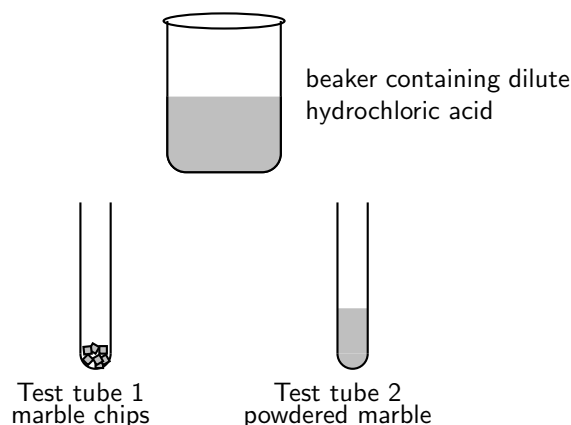


Aim:

To determine the effect of the surface area of reactants on the rate of the reaction.

Apparatus:

2 g marble chips, 2 g powdered marble, hydrochloric acid, beaker, two test tubes.



Method:

1. Prepare a solution of hydrochloric acid in the beaker by adding 2 cm^3 of the concentrated solution to 20 cm^3 of water.
2. Place the marble chips and powdered marble into separate test tubes.
3. Add 10 cm^3 of the dilute hydrochloric acid to each of the test tubes and observe the rate at which carbon dioxide gas is produced.

Results:

- Which reaction proceeds the fastest?
- Can you explain this?

Conclusions:

The reaction with powdered marble is the fastest. The smaller the pieces of marble are, the greater the surface area for the reaction to take place. The greater the surface area of the reactants, the faster the reaction rate will be.

Activity :: Experiment : Reactant concentration and reaction rate.**Aim:**

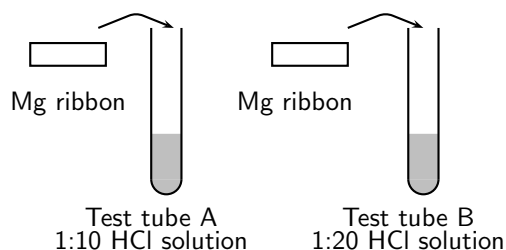
To determine the effect of reactant concentration on reaction rate.

Apparatus:

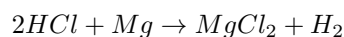
Concentrated hydrochloric acid (HCl), magnesium ribbon, two beakers, two test tubes, measuring cylinder.

Method:

1. Prepare a solution of dilute hydrochloric acid in one of the beakers by diluting 1 part concentrated acid with 10 parts water. For example, if you measure 1 cm³ of concentrated acid in a measuring cylinder and pour it into a beaker, you will need to add 10 cm³ of water to the beaker as well. In the same way, if you pour 2 cm³ of concentrated acid into a beaker, you will need to add 20 cm³ of water. Both of these are 1:10 solutions. Pour 10 cm³ of the 1:10 solution into a test tube and mark it 'A'. Remember to add the *acid* to the *water*, and not the other way around.
2. Prepare a second solution of dilute hydrochloric acid by diluting 1 part concentrated acid with 20 parts water. Pour 10cm³ of this 1:20 solution into a second test tube and mark it 'B'.
3. Take two pieces of magnesium ribbon of the **same length**. At the same time, put one piece of magnesium ribbon into test tube A and the other into test tube B, and observe closely what happens.



The equation for the reaction is:

**Results:**

- Which of the two solutions is more concentrated, the 1:10 or 1:20 hydrochloric acid solution?
- In which of the test tubes is the reaction the fastest? Suggest a reason for this.
- How can you measure the rate of this reaction?
- What is the gas that is given off?
- Why was it important that the same length of magnesium ribbon was used for each reaction?

Conclusions:

The 1:10 solution is more concentrated and this reaction therefore proceeds faster. The greater the concentration of the reactants, the faster the rate of the reaction. The rate of the reaction can be measured by the rate at which hydrogen gas is produced.

Activity :: Group work : The effect of temperature on reaction rate

1. In groups of 4-6, design an experiment that will help you to see the effect of temperature on the reaction time of 2 cm of magnesium ribbon and 20 ml of vinegar. During your group discussion, you should think about the following:
 - What equipment will you need?
 - How will you conduct the experiment to make sure that you are able to compare the results for different temperatures?
 - How will you record your results?
 - What safety precautions will you need to take when you carry out this experiment?
 2. Present your experiment ideas to the rest of the class, and give them a chance to comment on what you have done.
 3. Once you have received feedback, carry out the experiment and record your results.
 4. What can you conclude from your experiment?
-

3.3 Reaction rates and collision theory

It should be clear now that the rate of a reaction varies depending on a number of factors. But how can we *explain* why reactions take place at different speeds under different conditions? Why, for example, does an increase in the surface area of the reactants also increase the rate of the reaction? One way to explain this is to use **collision theory**.

For a reaction to occur, the particles that are reacting must collide with one another. Only a fraction of all the collisions that take place actually cause a chemical change. These are called 'successful' collisions. When there is an increase in the *concentration* of reactants, the chance that reactant particles will collide with each other also increases because there are more particles in that space. In other words, the *collision frequency* of the reactants increases. The number of *successful* collisions will therefore also increase, and so will the rate of the reaction. In the same way, if the *surface area* of the reactants increases, there is also a greater chance that successful collisions will occur.

**Definition: Collision theory**

Collision theory is a theory that explains how chemical reactions occur and why reaction rates differ for different reactions. The theory states that for a reaction to occur the reactant particles must collide, but that only a certain fraction of the total collisions, the *effective collisions*, actually cause the reactant molecules to change into products. This is because only a small number of the molecules have enough energy and the right orientation at the moment of impact to break the existing bonds and form new bonds.

When the *temperature* of the reaction increases, the average kinetic energy of the reactant particles increases and they will move around much more actively. They are therefore more likely to collide with one another (Figure 3.1). Increasing the temperature also increases the number of particles whose energy will be greater than the activation energy for the reaction (refer section 3.5).

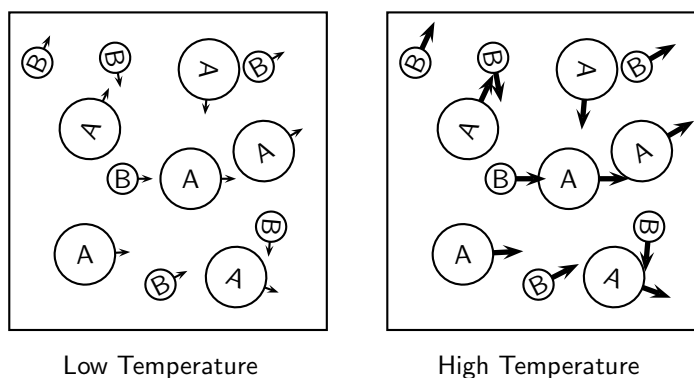
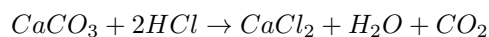


Figure 3.1: An increase in the temperature of a reaction increases the chances that the reactant particles (A and B) will collide because the particles have more energy and move around more.



Exercise: Rates of reaction

Hydrochloric acid and calcium carbonate react according to the following equation:



The volume of carbon dioxide that is produced during the reaction is measured at different times. The results are shown in the table below.

Time (mins)	Total Volume of CO ₂ produced (cm ³)
1	14
2	26
3	36
4	44
5	50
6	58
7	65
8	70
9	74
10	77

Note: On a graph of production against time, it is the gradient of the graph that shows the rate of the reaction.

Questions:

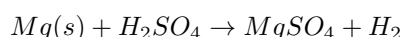
- Use the data in the table to draw a graph showing the volume of gas that is produced in the reaction, over a period of 10 minutes.
- At which of the following times is the reaction *fastest*? Time = 1 minute; time = 6 minutes or time = 8 minutes.
- Suggest a reason why the reaction slows down over time.
- Use the graph to estimate the volume of gas that will have been produced after 11 minutes.
- After what time do you think the reaction will stop?
- If the experiment was repeated using a more concentrated hydrochloric acid solution...
 - would the rate of the reaction increase or decrease from the one shown in the graph?
 - draw a rough line on the graph to show how you would expect the reaction to proceed with a more concentrated HCl solution.

3.4 Measuring Rates of Reaction

How the rate of a reaction is measured will depend on what the reaction is, and what product forms. Look back to the reactions that have been discussed so far. In each case, how was the rate of the reaction measured? The following examples will give you some ideas about other ways to measure the rate of a reaction:

- *Reactions that produce hydrogen gas:*

When a metal dissolves in an acid, hydrogen gas is produced. A lit splint can be used to test for hydrogen. The 'pop' sound shows that hydrogen is present. For example, magnesium reacts with sulfuric acid to produce magnesium sulphate and hydrogen.



- *Reactions that produce carbon dioxide:*

When a carbonate dissolves in an acid, carbon dioxide gas is produced. When carbon dioxide is passed through limewater, it turns the limewater milky. This is a simple test for the presence of carbon dioxide. For example, calcium carbonate reacts with hydrochloric acid to produce calcium chloride, water and carbon dioxide.



- *Reactions that produce gases such as oxygen or carbon dioxide:*

Hydrogen peroxide decomposes to produce oxygen. The volume of oxygen produced can be measured using the gas syringe method (figure 3.2). The gas collects in the syringe, pushing out against the plunger. The volume of gas that has been produced can be read from the markings on the syringe. For example, hydrogen peroxide decomposes in the presence of a manganese(IV) oxide catalyst to produce oxygen and water.

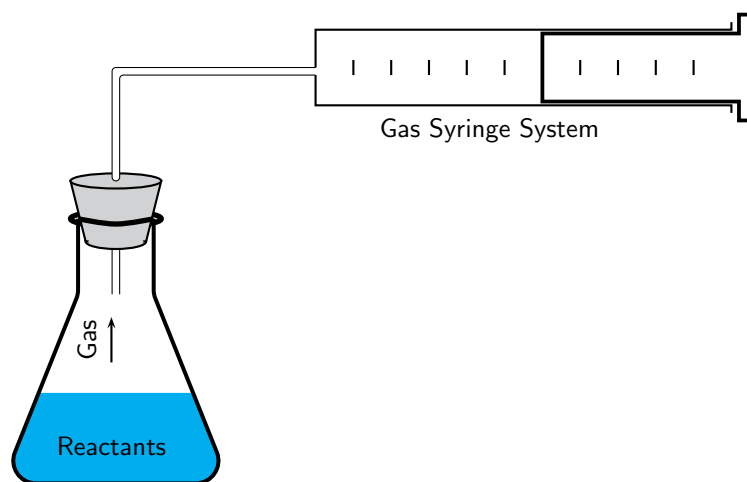
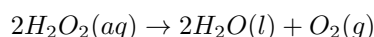


Figure 3.2: Gas Syringe Method

- *Precipitate reactions:*

In reactions where a *precipitate* is formed, the amount of precipitate formed in a period of time can be used as a measure of the reaction rate. For example, when sodium thiosulphate reacts with an acid, a yellow precipitate of sulfur is formed. The reaction is as follows:



One way to estimate the rate of this reaction is to carry out the investigation in a conical flask and to place a piece of paper with a black cross underneath the bottom of the flask. At the beginning of the reaction, the cross will be clearly visible when you look into the flask (figure 3.3). However, as the reaction progresses and more precipitate is formed, the cross will gradually become less clear and will eventually disappear altogether. Noting the time that it takes for this to happen will give an idea of the reaction rate. Note that it is not possible to collect the SO_2 gas that is produced in the reaction, because it is very soluble in water.

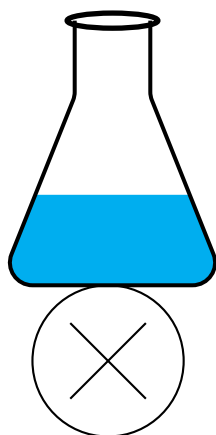


Figure 3.3: At the beginning of the reaction between sodium thiosulphate and hydrochloric acid, when no precipitate has been formed, the cross at the bottom of the conical flask can be clearly seen.

- *Changes in mass:*

The rate of a reaction that produces a gas can also be measured by calculating the mass loss as the gas is formed and escapes from the reaction flask. This method can be used for reactions that produce carbon dioxide or oxygen, but are not very accurate for reactions that give off hydrogen because the mass is too low for accuracy. Measuring changes in mass may also be suitable for other types of reactions.

Activity :: Experiment : Measuring reaction rates

Aim:

To measure the effect of concentration on the rate of a reaction.

Apparatus:

- 300 cm³ of sodium thiosulphate ($\text{Na}_2\text{S}_2\text{O}_3$) solution. Prepare a solution of sodium thiosulphate by adding 12 g of $\text{Na}_2\text{S}_2\text{O}_3$ to 300 cm³ of water. This is solution 'A'.
- 300 cm³ of water
- 100 cm³ of 1:10 dilute hydrochloric acid. This is solution 'B'.
- Six 100 cm³ glass beakers
- Measuring cylinders
- Paper and marking pen
- Stopwatch or timer

Method:

One way to measure the rate of this reaction is to place a piece of paper with a cross underneath the reaction beaker to see how quickly the cross is made invisible by the formation of the sulfur precipitate.

1. Set up six beakers on a flat surface and mark them from 1 to 6. Under each beaker you will need to place a piece of paper with a large black cross.
2. Pour 60 cm³ solution A into the first beaker and add 20 cm³ of water
3. Use the measuring cylinder to measure 10 cm³ HCl. Now add this HCl to the solution that is already in the first beaker (NB: Make sure that you always clean out the measuring cylinder you have used before using it for another chemical).
4. Using a stopwatch with seconds, record the time it takes for the precipitate that forms to block out the cross.
5. Now measure 50 cm³ of solution A into the second beaker and add 30 cm³ of water. To this second beaker, add 10 cm³ HCl, time the reaction and record the results as you did before.
6. Continue the experiment by diluting solution A as shown below.

Beaker	Solution A (cm ³)	Water (cm ³)	Solution B (cm ³)	Time (s)
1	60	20	10	
2	50	30	10	
3	40	40	10	
4	30	50	10	
5	20	60	10	
6	10	70	10	

The equation for the reaction between sodium thiosulphate and hydrochloric acid is:



Results:

- Calculate the reaction rate in each beaker. This can be done using the following equation:

$$\text{Rate of reaction} = \frac{1}{\text{time}}$$

- Represent your results on a graph. **Concentration** will be on the x-axis and **reaction rate** on the y-axis. Note that the original volume of Na₂S₂O₃ can be used as a measure of concentration.
- Why was it important to keep the volume of HCl constant?
- Describe the relationship between concentration and reaction rate.

Conclusions:

The rate of the reaction is fastest when the concentration of the reactants was the highest.

3.5 Mechanism of reaction and catalysis

Earlier it was mentioned that it is the *collision* of particles that causes reactions to occur and that only some of these collisions are 'successful'. This is because the reactant particles have a wide range of kinetic energy, and only a small fraction of the particles will have enough energy to actually break bonds so that a chemical reaction can take place. The minimum energy that is needed for a reaction to take place is called the **activation energy**. For more information on the energy of reactions, refer to Grade 11.

**Definition: Activation energy**

The energy that is needed to break the bonds in reactant molecules so that a chemical reaction can proceed.

Even at a fixed temperature, the energy of the particles varies, meaning that only some of them will have enough energy to be part of the chemical reaction, depending on the activation energy for that reaction. This is shown in figure 3.4. Increasing the reaction temperature has the effect of increasing the number of particles with enough energy to take part in the reaction, and so the reaction rate increases.

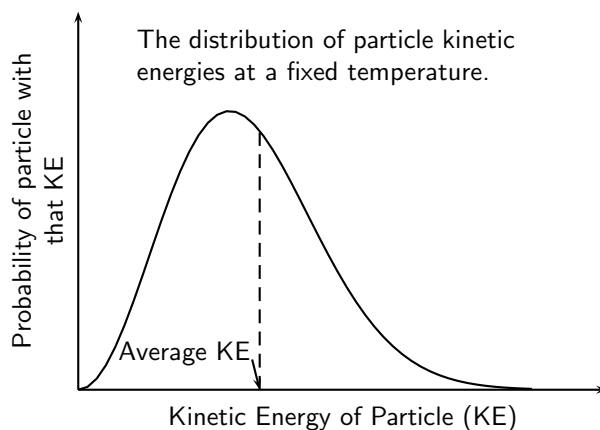
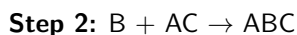
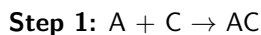


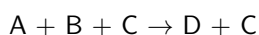
Figure 3.4: The distribution of particle kinetic energies at a fixed temperature

A **catalyst** functions slightly differently. The function of a catalyst is to lower the activation energy so that more particles now have enough energy to react. The catalyst itself is not changed during the reaction, but simply provides an alternative pathway for the reaction, so that it needs less energy. Some *metals* e.g. platinum, copper and iron can act as catalysts in certain reactions. In our own human bodies, *enzymes* are catalysts that help to speed up biological reactions. Catalysts generally react with one or more of the reactants to form a chemical intermediate which then reacts to form the final product. The chemical intermediate is sometimes called the **activated complex**.

The following is an example of how a reaction that involves a catalyst might proceed. C represents the catalyst, A and B are reactants and D is the product of the reaction of A and B.



In the above, ABC represents the intermediate chemical. Although the catalyst (C) is consumed by reaction 1, it is later produced again by reaction 4, so that the overall reaction is as follows:



You can see from this that the catalyst is released at the end of the reaction, completely unchanged.

**Definition: Catalyst**

A catalyst speeds up a chemical reaction, without being altered in any way. It increases the reaction rate by lowering the activation energy for a reaction.

Energy diagrams are useful to illustrate the effect of a **catalyst** on reaction rates. Catalysts decrease the activation energy required for a reaction to proceed (shown by the smaller 'hump' on the energy diagram in figure 3.5), and therefore increase the reaction rate.

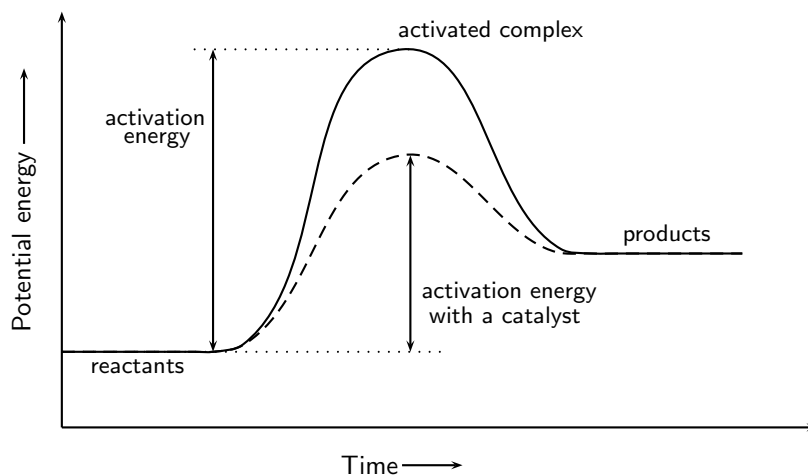


Figure 3.5: The effect of a catalyst on the activation energy of a reaction

Activity :: Experiment : Catalysts and reaction rates

Aim:

To determine the effect of a catalyst on the rate of a reaction

Apparatus:

Zinc granules, 0.1 M hydrochloric acid, copper pieces, one test tube and a glass beaker.

Method:

1. Place a few of the zinc granules in the test tube.
2. Measure the mass of a few pieces of copper and keep them separate from the rest of the copper.
3. Add about 20 cm³ of HCl to the test tube. You will see that a gas is released. Take note of how quickly or slowly this gas is released. Write a balanced equation for the chemical reaction that takes place.
4. Now add the copper pieces to the same test tube. What happens to the rate at which the gas is produced?
5. Carefully remove the copper pieces from the test tube (do not get HCl on your hands), rinse them in water and alcohol and then weigh them again. Has the mass of the copper changed since the start of the experiment?

Results:

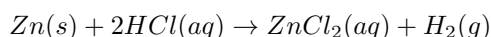
During the reaction, the gas that is released is hydrogen. The rate at which the hydrogen is produced *increases* when the copper pieces (the catalyst) are added. The mass of the copper does not change during the reaction.

Conclusions:

The copper acts as a *catalyst* during the reaction. It speeds up the rate of the reaction, but is not changed in any way itself.

**Exercise: Reaction rates**

- For each of the following, say whether the statement is **true** or **false**. If it is false, re-write the statement correctly.
 - A catalyst increases the energy of reactant molecules so that a chemical reaction can take place.
 - Increasing the temperature of a reaction has the effect of increasing the number of reactant particles that have more energy than the activation energy.
 - A catalyst does not become part of the final product in a chemical reaction.
- 5 g of zinc granules are added to 400 cm³ of 0.5 mol.dm⁻³ hydrochloric acid. To investigate the rate of the reaction, the change in the mass of the flask containing the zinc and the acid was measured by placing the flask on a direct reading balance. The reading on the balance shows that there is a decrease in mass during the reaction. The reaction which takes place is given by the following equation:



- Why is there a decrease in mass during the reaction?
 - The experiment is repeated, this time using 5 g of powdered zinc instead of granulated zinc. How will this influence the rate of the reaction?
 - The experiment is repeated once more, this time using 5 g of granulated zinc and 600 cm³ of 0.5 mol.dm⁻³ hydrochloric acid. How does the rate of this reaction compare to the **original** reaction rate?
 - What effect would a catalyst have on the rate of this reaction?
- (IEB Paper 2 2003)
- Enzymes are catalysts. Conduct your own research to find the names of common enzymes in the human body and which chemical reactions they play a role in.
 - 5 g of calcium carbonate powder reacts with 20 cm³ of a 0.1 mol.dm⁻³ solution of hydrochloric acid. The gas that is produced at a temperature of 25^{circ}C is collected in a gas syringe.
 - Write a balanced chemical equation for this reaction.
 - The rate of the reaction is determined by measuring the volume of gas that is produced in the first minute of the reaction. How would the rate of the reaction be affected if:
 - a lump of calcium carbonate of the same mass is used
 - 40 cm³ of 0.1 mol.dm⁻³ hydrochloric acid is used

3.6 Chemical equilibrium

Having looked at factors that affect the rate of a reaction, we now need to ask some important questions. Does a reaction always proceed in the same direction or can it be reversible? In other words, is it always true that a reaction proceeds from *reactants to products*, or is it possible that sometimes, the reaction will reverse and the *products will be changed back into the reactants*? And does a reaction always run its full course so that all the reactants are used up, or can a reaction reach a point where reactants are still present, but there does not seem to be any further change taking place in the reaction? The following demonstration might help to explain this.

Activity :: Demonstration : Liquid-vapour phase equilibrium**Apparatus and materials:**

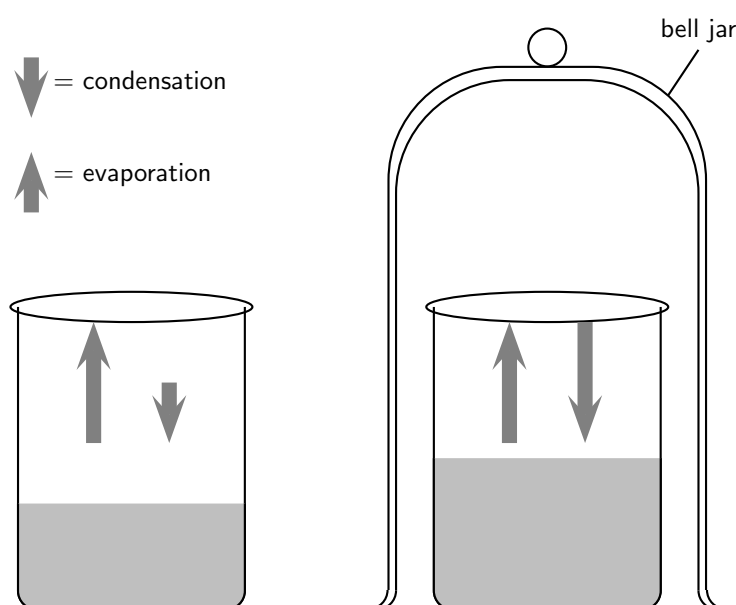
2 beakers; water; bell jar

Method:

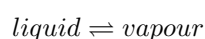
1. Half fill two beakers with water and mark the level of the water in each case.
2. Cover one of the beakers with a bell jar.
3. Leave the beakers and, over the course of a day or two, observe how the water level in the two beakers changes. What do you notice? Note: You could speed up this demonstration by placing the two beakers over a bunsen burner to heat the water. In this case, it may be easier to cover the second beaker with a glass cover.

Observations:

You should notice that in the beaker that is uncovered, the water level drops quickly because of evaporation. In the beaker that is covered, there is an initial drop in the water level, but after a while evaporation appears to stop and the water level in this beaker is higher than that in the one that is open. Note that the diagram below shows the situation at time=0.

**Discussion:**

In the first beaker, liquid water becomes water vapour as a result of evaporation and the water level drops. In the second beaker, evaporation also takes place. However, in this case, the vapour comes into contact with the surface of the bell jar and it cools and condenses to form liquid water again. This water is returned to the beaker. Once condensation has begun, the rate at which water is lost from the beaker will start to decrease. At some point, the rate of evaporation will be equal to the rate of condensation above the beaker, and there will be no change in the water level in the beaker. This can be represented as follows:



In this example, the reaction (in this case, a change in the phase of water) can proceed in either direction. In one direction there is a change in phase from liquid to vapour. But the reverse can also take place, when vapour condenses to form water again.

In a **closed system** it is possible for reactions to be reversible, such as in the demonstration above. In a closed system, it is also possible for a chemical reaction to reach **equilibrium**. We will discuss these concepts in more detail.

3.6.1 Open and closed systems

An **open system** is one in which matter or energy can flow into or out of the system. In the liquid-vapour demonstration we used, the first beaker was an example of an open system because the beaker could be heated or cooled (a change in *energy*), and water vapour (the *matter*) could evaporate from the beaker.

A **closed system** is one in which energy can enter or leave, but matter cannot. The second beaker covered by the bell jar is an example of a closed system. The beaker can still be heated or cooled, but water vapour cannot leave the system because the bell jar is a barrier. Condensation changes the vapour to liquid and returns it to the beaker. In other words, there is no loss of matter from the system.

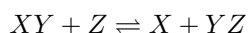


Definition: Open and closed systems

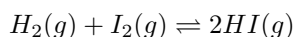
An *open system* is one whose borders allow the movement of energy and matter into and out of the system. A *closed system* is one in which only energy can be exchanged, but not matter.

3.6.2 Reversible reactions

Some reactions can take place in two directions. In one direction the reactants combine to form the products. This is called the **forward reaction**. In the other, the products react to form reactants again. This is called the **reverse reaction**. A special double-headed arrow is used to show this type of **reversible reaction**:



So, in the following reversible reaction:



The forward reaction is $H_2(g) + I_2(g) \rightarrow 2HI(g)$. The reverse reaction is $2HI(g) \rightarrow H_2(g) + I_2(g)$.



Definition: A reversible reaction

A reversible reaction is a chemical reaction that can proceed in both the forward and reverse directions. In other words, the reactant and product of one reaction may reverse roles.

Activity :: Demonstration : The reversibility of chemical reactions

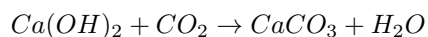
Apparatus and materials:

Lime water ($Ca(OH)_2$); calcium carbonate ($CaCO_3$); hydrochloric acid; 2 test tubes with rubber stoppers; delivery tube; retort stand and clamp; bunsen burner.

Method and observations:

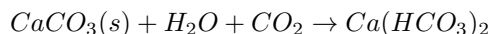
1. Half-fill a test tube with clear lime water ($Ca(OH)_2$).
2. In another test tube, place a few pieces of calcium carbonate ($CaCO_3$) and cover the pieces with dilute hydrochloric acid. Seal the test tube with a rubber stopper and delivery tube.
3. Place the other end of the delivery tube into the test tube containing the lime water so that the carbon dioxide that is produced from the reaction between calcium carbonate and hydrochloric acid passes through the lime water. Observe what happens to the appearance of the lime water.

The equation for the reaction that takes place is:



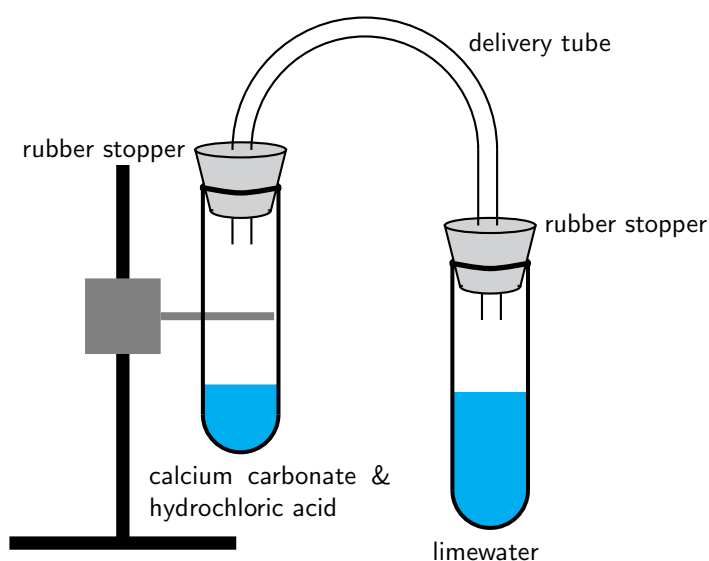
CaCO_3 is insoluble and it turns the limewater milky.

4. Allow the reaction to proceed for a while so that carbon dioxide continues to pass through the limewater. What do you notice? The equation for the reaction that takes place is:



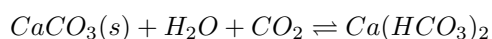
In this reaction, calcium carbonate becomes one of the reactants to produce hydrogen carbonate ($\text{Ca}(\text{HCO}_3)_2$) and so the solution becomes clear again.

5. Heat the solution in the test tube over a bunsen burner. What do you observe? You should see bubbles of carbon dioxide appear and the limewater turns milky again. The reaction that has taken place is:



Discussion:

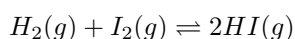
- If you look at the last two equations you will see that the one is the reverse of the other. In other words, this is a *reversible reaction* and can be written as follows:



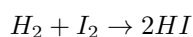
- Is the forward reaction endothermic or exothermic? Is the reverse reaction endothermic or exothermic? You should have noticed that the reverse reaction only took place when the solution was heated. Sometimes, changing the temperature of a reaction can change its direction.

3.6.3 Chemical equilibrium

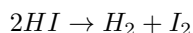
Using the same reversible reaction that we used in an earlier example:



The forward reaction is:



The reverse reaction is:



When the rate of the forward reaction and the reverse reaction are equal, the system is said to be in **equilibrium**. Figure 3.6 shows this. Initially (time = 0), the rate of the forward reaction is high and the rate of the reverse reaction is low. As the reaction proceeds, the rate of the forward reaction decreases and the rate of the reverse reaction increases, until both occur at the same rate. This is called equilibrium.

Although it is not always possible to observe any macroscopic changes, this does not mean that the reaction has stopped. The forward and reverse reactions continue to take place and so microscopic changes still occur in the system. This state is called **dynamic equilibrium**. In the liquid-vapour phase equilibrium demonstration, dynamic equilibrium was reached when there was no observable change in the level of the water in the second beaker even though evaporation and condensation continued to take place.

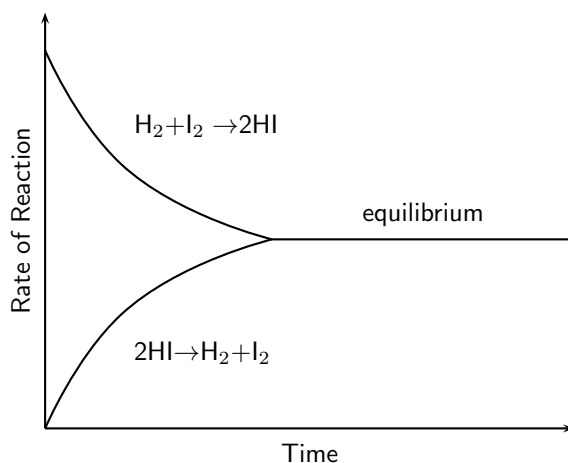


Figure 3.6: The change in rate of forward and reverse reactions in a closed system

There are, however, a number of factors that can change the chemical equilibrium of a reaction. Changing the **concentration**, the **temperature** or the **pressure** of a reaction can affect equilibrium. These factors will be discussed in more detail later in this chapter.



Definition: Chemical equilibrium

Chemical equilibrium is the state of a chemical reaction, where the concentrations of the reactants and products have no net change over time. Usually, this state results when the forward chemical reactions proceed at the same rate as their reverse reactions.

3.7 The equilibrium constant

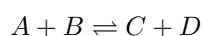


Definition: Equilibrium constant

The equilibrium constant (K_c), relates to a chemical reaction at equilibrium. It can be calculated if the equilibrium concentration of each reactant and product in a reaction at equilibrium is known.

3.7.1 Calculating the equilibrium constant

Consider the following generalised reaction which takes place in a **closed** container at a **constant temperature**:



We know from section 3.2 that the rate of the forward reaction is directly proportional to the concentration of the reactants. In other words, as the concentration of the reactants increases, so does the rate of the forward reaction. This can be shown using the following equation:

$$\text{Rate of forward reaction} \propto [A][B]$$

or

$$\text{Rate of forward reaction} = k_1[A][B]$$

Similarly, the rate of the reverse reaction is directly proportional to the concentration of the products. This can be shown using the following equation:

$$\text{Rate of reverse reaction} \propto [C][D]$$

or

$$\text{Rate of reverse reaction} = k_2[C][D]$$

At equilibrium, the rate of the forward reaction is equal to the rate of the reverse reaction. This can be shown using the following equation:

$$k_1[A][B] = k_2[C][D]$$

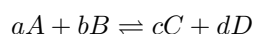
or

$$\frac{k_1}{k_2} = \frac{[C][D]}{[A][B]}$$

or, if the constants k_1 and k_2 are simplified to a single constant, the equation becomes:

$$k_c = \frac{[C][D]}{[A][B]}$$

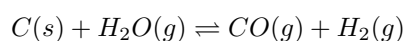
A more general form of the equation for a reaction at chemical equilibrium is:



where A and B are reactants, C and D are products and a, b, c, and d are the coefficients of the respective reactants and products. A more general formula for calculating the equilibrium constant is therefore:

$$K_c = \frac{[C]^c[D]^d}{[A]^a[B]^b}$$

It is important to note that if a reactant or a product in a chemical reaction is in either the **liquid** or **solid** phase, the concentration stays constant during the reaction. Therefore, these values can be left out of the equation to calculate K_c . For example, in the following reaction:



$$K_c = \frac{[CO][H_2]}{[H_2O]}$$

Important:

1. The constant K_c is affected by temperature and so, if the values of K_c are being compared for different reactions, it is important that all the reactions have taken place at the same temperature.
2. K_c values do not have units. If you look at the equation, the units all cancel each other out.

3.7.2 The meaning of K_c values

The formula for K_c has the concentration of the products in the numerator and the concentration of reactants in the denominator. So a *high* K_c value means that the concentration of products is high and the reaction has a high yield. We can also say that the equilibrium lies far to the right. The opposite is true for a low K_c value. A *low* K_c value means that, at equilibrium, there are more reactants than products and therefore the yield is low. The equilibrium for the reaction lies far to the left.

Important: Calculations made easy

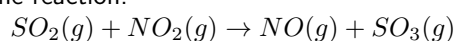
When you are busy with calculations that involve the equilibrium constant, the following tips may help:

1. Make sure that you always read the question carefully to be sure of what you are being asked to calculate. If the equilibrium constant is involved, make sure that the concentrations you use are the concentrations **at equilibrium**, and not the concentrations or quantities that are present at some other time in the reaction.
2. When you are doing more complicated calculations, it sometimes helps to draw up a table like the one below and fill in the **mole values** that you know or those you can calculate. This will give you a clear picture of what is happening in the reaction and will make sure that you use the right values in your calculations.

	Reactant 1	Reactant 2	Product 1
Start of reaction			
Used up			
Produced			
Equilibrium			

Worked Example 13: Calculating K_c

Question: For the reaction:



the concentration of the reagents is as follows:

$$[SO_3] = 0.2 \text{ mol.dm}^{-3}$$

$$[NO_2] = 0.1 \text{ mol.dm}^{-3}$$

$$[NO] = 0.4 \text{ mol.dm}^{-3}$$

$$[SO_2] = 0.2 \text{ mol.dm}^{-3}$$

Calculate the value of K_c .

Answer

Step 1 : Write the equation for k_c

$$K_c = \frac{[NO][SO_3]}{[SO_2][NO_2]}$$

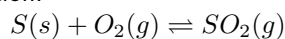
Step 2 : Fill in the values you know for this equation and calculate K_c

$$K_c = \frac{(0.4 \times 0.2)}{(0.2 \times 0.1)} = 4$$



Worked Example 14: Calculating reagent concentration

Question: For the reaction:



1. Write an equation for the equilibrium constant.
2. Calculate the equilibrium concentration of O_2 if $K_c=6$ and $[SO_2] = 3\text{mol.dm}^{-3}$ at equilibrium.

Answer

Step 3 : Write the equation for K_c

$$K_c = \frac{[SO_2]}{[O_2]}$$

(Sulfur is left out of the equation because it is a solid and its concentration stays constant during the reaction)

Step 4 : Re-arrange the equation so that oxygen is on its own on one side of the equation

$$[O_2] = \frac{[SO_2]}{K_c}$$

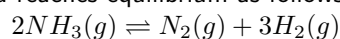
Step 5 : Fill in the values you know and calculate $[O_2]$

$$[O_2] = \frac{3\text{mol.dm}^{-3}}{6} = 0.5\text{mol.dm}^{-3}$$



Worked Example 15: Equilibrium calculations

Question: Initially 1.4 moles of $NH_3(g)$ is introduced into a sealed 2.0 dm^{-3} reaction vessel. The ammonia decomposes when the temperature is increased to 600K and reaches equilibrium as follows:



When the equilibrium mixture is analysed, the concentration of $\text{NH}_3(\text{g})$ is $0.3 \text{ mol}\cdot\text{dm}^{-3}$

1. Calculate the concentration of $\text{N}_2(\text{g})$ and $\text{H}_2(\text{g})$ in the equilibrium mixture.
2. Calculate the equilibrium constant for the reaction at 900 K.

Answer

Step 1 : Calculate the number of moles of NH_3 at equilibrium.

$$c = \frac{n}{V}$$

Therefore,

$$n = c \times V = 0.3 \times 2 = 0.6 \text{ mol}$$

Step 2 : Calculate the number of moles of ammonia that react (are 'used up') in the reaction.

$$\text{Moles used up} = 1.4 - 0.6 = 0.8 \text{ moles}$$

Step 3 : Calculate the number of moles of product that are formed.

Remember to use the mole ratio of reactants to products to do this. In this case, the ratio of $\text{NH}_3:\text{N}_2:\text{H}_2 = 2:1:3$. Therefore, if 0.8 moles of ammonia are used up in the reaction, then 0.4 moles of nitrogen are produced and 1.2 moles of hydrogen are produced.

Step 4 : Complete the following table

	NH_3	N_2	H_2
Start of reaction	1.4	0	0
Used up	0.8	0	0
Produced	0	0.4	1.2
Equilibrium	0.6	0.4	1.2

Step 5 : Using the values in the table, calculate $[\text{N}_2]$ and $[\text{H}_2]$

$$[\text{N}_2] = \frac{n}{V} = \frac{0.4}{2} = 0.2 \text{ mol}\cdot\text{dm}^{-3}$$

$$[\text{H}_2] = \frac{n}{V} = \frac{1.2}{2} = 0.6 \text{ mol}\cdot\text{dm}^{-3}$$

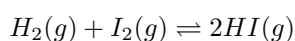
Step 6 : Calculate K_c

$$K_c = \frac{[\text{H}_2]^3[\text{N}_2]}{[\text{NH}_3]^2} = \frac{(0.6)^3(0.2)}{(0.3)^2} = 0.48$$



Worked Example 16: Calculating K_c

Question: Hydrogen and iodine gas react according to the following equation:



When 0.496 mol H_2 and 0.181 mol I_2 are heated at 450°C in a 1 dm^3

container, the equilibrium mixture is found to contain 0.00749 mol I_2 . Calculate the equilibrium constant for the reaction at 450°C.

Answer

Step 1 : Calculate the number of moles of iodine used in the reaction.

Moles of iodine used = 0.181 - 0.00749 = 0.1735 mol

Step 2 : Calculate the number of moles of hydrogen that are used up in the reaction.

The mole ratio of hydrogen:iodine = 1:1, therefore 0.1735 moles of hydrogen must also be used up in the reaction.

Step 3 : Calculate the number of moles of hydrogen iodide that are produced.

The mole ratio of $H_2:I_2:HI$ = 1:1:2, therefore the number of moles of HI produced is $0.1735 \times 2 = 0.347$ mol.

So far, the table can be filled in as follows:

	H_2 (g)	I_2	$2HI$
Start of reaction	0.496	0.181	0
Used up	0.1735	0.1735	0
Produced	0	0	0.347
Equilibrium	0.3225	0.0075	0.347

Step 4 : Calculate the concentration of each of the reactants and products at equilibrium.

$$c = \frac{n}{V}$$

Therefore the equilibrium concentrations are as follows:

$$[H_2] = 0.3225 \text{ mol.dm}^{-3}$$

$$[I_2] = 0.0075 \text{ mol.dm}^{-3}$$

$$[HI] = 0.347 \text{ mol.dm}^{-3}$$

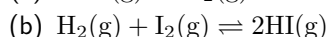
Step 5 : Calculate K_c

$$K_c = \frac{[HI]}{[H_2][I_2]} = \frac{0.347}{0.3225 \times 0.0075} = 143.47$$

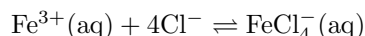


Exercise: The equilibrium constant

1. Write the equilibrium constant expression, K_c for the following reactions:

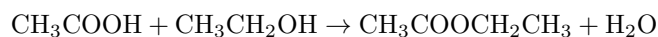


2. The following reaction takes place:



K_c for the reaction is $7.5 \times 10^{-2} \text{ mol.dm}^{-3}$. At equilibrium, the concentration of $FeCl_4^-$ is $0.95 \times 10^{-4} \text{ mol.dm}^{-3}$ and the concentration of free iron (Fe^{3+}) is 0.2 mol.dm^{-3} . Calculate the concentration of chloride ions at equilibrium.

3. Ethanoic acid (CH_3COOH) reacts with ethanol (CH_3CH_2OH) to produce ethyl ethanoate and water. The reaction is:



At the beginning of the reaction, there are 0.5 mols of ethanoic acid and 0.5 mols of ethanol. At equilibrium, 0.3 mols of ethanoic acid was left unreacted. The volume of the reaction container is 2 dm³. Calculate the value of K_c .

3.8 Le Chatelier's principle

A number of factors can influence the equilibrium of a reaction. These are:

1. concentration
2. temperature
3. pressure

Le Chatelier's Principle helps to predict what a change in temperature, concentration or pressure will have on the position of the equilibrium in a chemical reaction. This is very important, particularly in industrial applications, where yields must be accurately predicted and maximised.

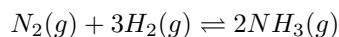


Definition: Le Chatelier's Principle

If a chemical system at equilibrium experiences a change in concentration, temperature or total pressure the equilibrium will shift in order to minimise that change and a new equilibrium is established.

3.8.1 The effect of concentration on equilibrium

If the concentration of a substance is increased, the equilibrium will shift so that this concentration decreases. So for example, if the concentration of a reactant was increased, the equilibrium would shift in the direction of the reaction that *uses up* the reactants, so that the reactant concentration decreases and equilibrium is restored. In the reaction between nitrogen and hydrogen to produce ammonia:

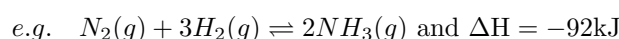


- If the nitrogen or hydrogen concentration was increased, Le Chatelier's principle predicts that equilibrium will shift to favour the *forward reaction* so that the excess nitrogen and hydrogen are used up to produce ammonia. Equilibrium shifts to the *right*.
- If the nitrogen or hydrogen concentration was decreased, the *reverse reaction* would be favoured so that some of the ammonia would change back to nitrogen and hydrogen to restore equilibrium.
- The same would be true if the concentration of the product (NH₃) was changed. If [NH₃] decreases, the forward reaction is favoured and if [NH₃] increases, the reverse reaction is favoured.

3.8.2 The effect of temperature on equilibrium

If the temperature of a reaction mixture is increased, the equilibrium will shift to decrease the temperature. So it will favour the reaction which will *use up* heat energy, in other words the endothermic reaction. The opposite is true if the temperature is decreased. In this case, the reaction that *produces* heat energy will be favoured, in other words, the exothermic reaction.

The reaction shown below is exothermic (shown by the negative value for ΔH). This means that the forward reaction, where nitrogen and hydrogen react to form ammonia, gives off heat. In the reverse reaction, where ammonia is broken down into hydrogen and nitrogen gas, heat is used up and so this reaction is endothermic.



An increase in temperature favours the reaction that is endothermic (the reverse reaction) because it uses up energy. If the temperature is increased, then the yield of ammonia (NH_3) decreases.

A decrease in temperature favours the reaction that is exothermic (the forward reaction) because it produces energy. Therefore, if the temperature is decreased, then the yield of NH_3 increases.

Activity :: Experiment : Le Chatelier's Principle

Aim:

To determine the effect of a change in concentration and temperature on chemical equilibrium

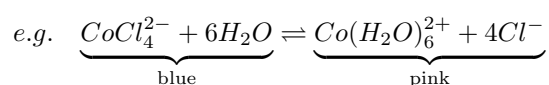
Apparatus:

0.2 M $CoCl_2$ solution, concentrated HCl, water, test tube, bunsen burner

Method:

1. Put 4-5 drops of 0.2M $CoCl_2$ solution into a test tube.
2. Add 20-25 drops of concentrated HCl.
3. Add 10-12 drops of water.
4. Heat the solution for 1-2 minutes.
5. Cool the solution for 1 minute under a tap.
6. Observe and record the colour changes that take place during the reaction.

The equation for the reaction that takes place is:



Results:

Complete your observations in the table below, showing the colour changes that take place, and also indicating whether the concentration of each of the ions in solution increases or decreases.

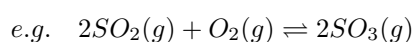
	Initial colour	Final colour	[Co ²⁺]	[Cl ⁻]	[CoCl ₄ ²⁻]
Add Cl ⁻					
Add H ₂ O					
Increase temp.					
Decrease temp.					

Conclusions:

Use your knowledge of equilibrium principles to explain the changes that you recorded in the table above. Draw a conclusion about the effect of a change in concentration of either the reactants or products on the equilibrium position. Also draw a conclusion about the effect of a change in temperature on the equilibrium position.

3.8.3 The effect of pressure on equilibrium

In the case of gases, we refer to pressure instead of concentration. Similar principles apply as those that were described before for concentration. When the pressure of a system increases, there are more particles in a particular space. The equilibrium will shift in a direction that reduces the number of gas particles so that the pressure is also reduced. To predict what will happen in a reaction, we need to look at the number of moles of gas that are in the reactants and products. Look at the example below:



In this reaction, two moles of product are formed for every three moles of reactants. If we increase the pressure on the closed system, the equilibrium will shift to the right because the forward reaction reduces the number of moles of gas that are present. This means that the yield of SO₃ will increase. The opposite will apply if the pressure on the system decreases. the equilibrium will shift to the left, and the concentration of SO₂ and O₂ will increase.



Important: The following rules will help in predicting the changes that take place in equilibrium reactions:

1. If the forward reaction that forms the product is endothermic, then an increase in temperature will favour this reaction and the yield of product will increase. Lowering the temperature will decrease the product yield.
2. If the forward reaction that forms the product is exothermic, then a decrease in temperature will favour this reaction and the product yield will increase. Increasing the temperature will decrease the product yield.
3. Increasing the pressure favours the side of the equilibrium with the least number of gas molecules. This is shown in the balanced symbol equation. This rule applies in reactions with one or more gaseous reactants or products.
4. Decreasing the pressure favours the side of the equilibrium with the most number of gas molecules. This rule applies in reactions with one or more gaseous reactants or products.
5. If the concentration of a reactant (on the left) is increased, then some of it must change to the products (on the right) for equilibrium to be maintained. The equilibrium position will shift to the right.
6. If the concentration of a reactant (on the left) is decreased, then some of the products (on the right) must change back to reactants for equilibrium to be maintained. The equilibrium position will shift to the left.
7. A **catalyst** does not affect the equilibrium position of a reaction. It only influences the *rate of the reaction*, in other words, how quickly equilibrium is reached.



Worked Example 17: Reaction Rates 1

Question: $2\text{NO}_2(g) \rightleftharpoons 2\text{NO}(g) + \text{O}_2(g)$ and $\Delta H > 0$ How will the rate of the reverse reaction be affected by:

1. a decrease in temperature?
2. the addition of a catalyst?
3. the addition of more NO gas?

Answer

1. The rate of the forward reaction will increase since it is the forward reaction that is exothermic and therefore *produces* energy to balance the loss of energy from the decrease in temperature. The rate of the reverse reaction will decrease.
2. The rate of the reverse and the forward reaction will increase.
3. The rate of the reverse reaction will increase so that the extra NO gas is converted into NO_2 gas.



Worked Example 18: Reaction Rates 2**Question:**

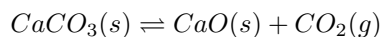
1. Write a balanced equation for the exothermic reaction between Zn(s) and HCl.
2. Name 3 ways to increase the reaction rate between hydrochloric acid and zinc metal.

Answer

1. $Zn(s) + 2HCl(aq) \rightleftharpoons ZnCl_2(aq) + H_2(g)$
2. A catalyst could be added, the zinc solid could be ground into a fine powder to increase its surface area, the HCl concentration could be increased or the reaction temperature could be increased.

**Exercise: Reaction rates and equilibrium**

1. The following reaction reaches equilibrium in a closed container:



The pressure of the system is increased by decreasing the volume of the container. How will the number of moles and the concentration of the $CO_2(g)$ have changed when a new equilibrium is reached at the same temperature?

	moles of CO_2	$[CO_2]$
A	decreased	decreased
B	increased	increased
C	decreased	stays the same
D	decreased	increased

(IEB Paper 2, 2003)

2. The following reaction has reached equilibrium in a closed container:

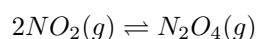


The pressure of the system is then decreased by increasing the volume of the container. How will the concentration of the $H_2(g)$ and the value of K_c be affected when the new equilibrium is established? Assume that the temperature of the system remains unchanged.

	$[H_2]$	K_c
A	increases	increases
B	increases	unchanged
C	unchanged	unchanged
D	decreases	unchanged

(IEB Paper 2, 2004)

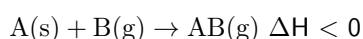
3. During a classroom experiment copper metal reacts with concentrated nitric acid to produce NO_2 gas, which is collected in a gas syringe. When enough gas has collected in the syringe, the delivery tube is clamped so that no gas can escape. The brown NO_2 gas collected reaches an equilibrium with colourless N_2O_4 gas as represented by the following equation:



Once this equilibrium has been established, there are 0.01 moles of NO_2 gas and 0.03 moles of N_2O_4 gas present in the syringe.

- (a) A learner, noticing that the colour of the gas mixture in the syringe is no longer changing, comments that all chemical reactions in the syringe must have stopped. Is this assumption correct? Explain.
- (b) The gas in the syringe is cooled. The volume of the gas is kept constant during the cooling process. Will the gas be lighter or darker at the lower temperature? Explain your answer.
- (c) The volume of the syringe is now reduced to 75 cm^3 by pushing the plunger in and holding it in the new position. There are 0.032 moles of N_2O_4 gas present once the equilibrium has been re-established at the reduced volume (75 cm^3). Calculate the value of the equilibrium constant for this equilibrium.
- (IEB Paper 2, 2004)

4. Consider the following reaction, which takes place in a closed container:

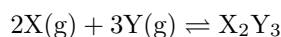


If you wanted to increase the rate of the reaction, which of the following would you do?

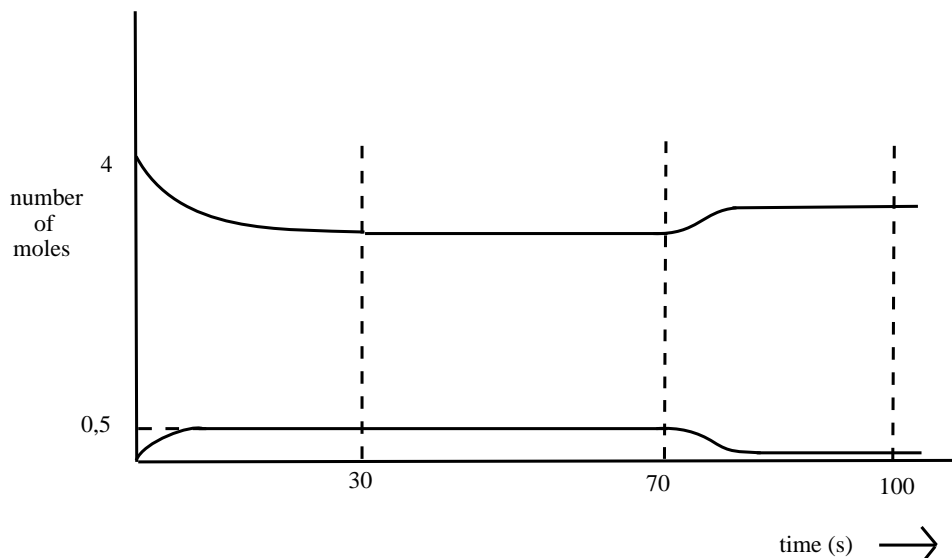
- (a) decrease the concentration of B
 (b) decrease the temperature of A
 (c) grind A into a fine powder
 (d) decrease the pressure

(IEB Paper 2, 2002)

5. Gases X and Y are pumped into a 2 dm^3 container. When the container is sealed, 4 moles of gas X and 4 moles of gas Y are present. The following equilibrium is established:



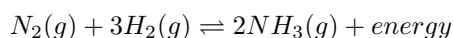
The graph below shows the number of moles of gas X and gas X_2Y_3 that are present from the time the container is sealed.



- (a) How many moles of gas X_2Y_3 are formed by the time the reaction reaches equilibrium at 30 seconds?
- (b) Calculate the value of the equilibrium constant at $t = 50 \text{ s}$.
- (c) At 70 s the temperature is increased. Is the forward reaction endothermic or exothermic? Explain in terms of Le Chatelier's Principle.
- (d) How will this increase in temperature affect the value of the equilibrium constant?

3.9 Industrial applications

The **Haber process** is a good example of an industrial process which uses the equilibrium principles that have been discussed. The equation for the process is as follows:



Since the reaction is **exothermic**, the forward reaction is favoured at low temperatures, and the reverse reaction at high temperatures. If the purpose of the Haber process is to produce ammonia, then the temperature must be maintained at a level that is low enough to ensure that the reaction continues in the forward direction.

The forward reaction is also favoured by high *pressures* because there are four moles of reactant for every two moles of product formed.

The K value for this reaction will be calculated as follows:

$$K = \frac{[NH_3]^2}{[N_2][H_2]^3}$$



Exercise: Applying equilibrium principles

Look at the values of k calculated for the Haber process reaction at different temperatures, and then answer the questions that follow:

$T^{\circ}C$	k
25	6.4×10^2
200	4.4×10^{-1}
300	4.3×10^{-3}
400	1.6×10^{-4}
500	1.5×10^{-5}

1. What happens to the value of K as the temperature increases?
2. Which reaction is being favoured when the temperature is 300 degrees celsius?
3. According to this table, which temperature would be best if you wanted to produce as much ammonia as possible? Explain.

3.10 Summary

- The **rate of a reaction** describes how quickly reactants are used up, or how quickly products form. The units used are moles per second.
- A number of factors can affect the rate of a reaction. These include the **nature of the reactants**, the **concentration** of reactants, **temperature** of the reaction, the presence or absence of a **catalyst** and the **surface area** of the reactants.
- **Collision theory** provides one way of explaining why each of these factors can affect the rate of a reaction. For example, higher temperatures mean increased reaction rates because the reactant particles have more energy and are more likely to collide successfully with each other.

- Different methods can be used to **measure the rate of a reaction**. The method used will depend on the nature of the product. Reactions that produce gases can be measured by collecting the gas in a syringe. Reactions that produce a precipitate are also easy to measure because the precipitate is easily visible.
- For any reaction to occur, a minimum amount of energy is needed so that bonds in the reactants can break, and new bonds can form in the products. The minimum energy that is required is called the **activation energy** of a reaction.
- In reactions where the particles do not have enough energy to overcome this activation energy, one of two methods can be used to facilitate a reaction to take place: increase the temperature of the reaction or add a catalyst.
- **Increasing the temperature of a reaction** means that the average energy of the reactant particles increases and they are more likely to have enough energy to overcome the activation energy.
- A **catalyst** is used to lower the activation energy so that the reaction is more likely to take place. A catalyst does this by providing an alternative, lower energy pathway, for the reaction.
- A catalyst therefore **speeds up a reaction** but does not become part of the reaction in any way.
- **Chemical equilibrium** is the state of a reaction, where the concentrations of the reactants and the products have no net change over time. Usually this occurs when the rate of the forward reaction is the same as the rate of the reverse reaction.
- The **equilibrium constant** relates to reactions at equilibrium, and can be calculated using the following equation:

$$K_c = \frac{[C]^c[D]^d}{[A]^a[B]^b}$$

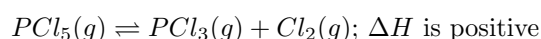
where A and B are reactants, C and D are products and a, b, c, and d are the coefficients of the respective reactants and products.

- A **high K_c value** means that the concentration of products at equilibrium is high and the reaction has a high yield. A **low K_c value** means that the concentration of products at equilibrium is low and the reaction has a low yield.
- **Le Chatelier's Principle** states that if a chemical system at equilibrium experiences a change in concentration, temperature or total pressure the equilibrium will shift in order to minimise that change and to re-establish equilibrium. For example, if the pressure of a gaseous system at equilibrium was increased, the equilibrium would shift to favour the reaction that produces the lowest quantity of the gas. If the temperature of the same system was to increase, the equilibrium would shift to favour the endothermic reaction. Similar principles apply for changes in concentration of the reactants or products in a reaction.
- The principles of equilibrium are very important in **industrial applications** such as the Haber process, so that productivity can be maximised.



Exercise: Summary Exercise

1. For each of the following questions, choose the one correct answer from the list provided.
 - (a) Consider the following reaction that has reached equilibrium after some time in a sealed 1 dm³ flask:



Which one of the following reaction conditions applied to the system would decrease the rate of the reverse reaction?

- increase the pressure
- increase the reaction temperature
- continually remove $\text{Cl}_2(\text{g})$ from the flask
- addition of a suitable catalyst

(IEB Paper 2, 2001)

- (b) The following equilibrium constant expression is given for a particular reaction:

$$K_c = [\text{H}_2\text{O}]^4[\text{CO}_2]^3/[\text{C}_3\text{H}_8][\text{O}_2]^5$$

For which one of the following reactions is the above expression of K_c is correct?

- $\text{C}_3\text{H}_8(\text{g}) + 5\text{O}_2(\text{g}) \rightleftharpoons 4\text{H}_2\text{O}(\text{g}) + 3\text{CO}_2(\text{g})$
- $4\text{H}_2\text{O}(\text{g}) + 3\text{CO}_2(\text{g}) \rightleftharpoons \text{C}_3\text{H}_8(\text{g}) + 5\text{O}_2(\text{g})$
- $2\text{C}_3\text{H}_8(\text{g}) + 7\text{O}_2(\text{g}) \rightleftharpoons 6\text{CO}(\text{g}) + 8\text{H}_2\text{O}(\text{g})$
- $\text{C}_3\text{H}_8(\text{g}) + 5\text{O}_2(\text{g}) \rightleftharpoons 4\text{H}_2\text{O}(\text{l}) + 3\text{CO}_2(\text{g})$

(IEB Paper 2, 2001)

2. 10 g of magnesium ribbon reacts with a $0.15 \text{ mol}\cdot\text{dm}^{-3}$ solution of hydrochloric acid at a temperature of 25°C .

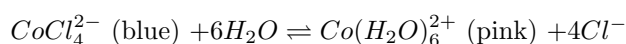
- Write a balanced chemical equation for the reaction.
- State **two** ways of increasing the rate of production of $\text{H}_2(\text{g})$.
- A table of the results is given below:

Time elapsed (min)	Vol of $\text{H}_2(\text{g})$ (cm^3)
0	0
0.5	17
1.0	25
1.5	30
2.0	33
2.5	35
3.0	35

- Plot a graph of volume versus time for these results.
- Explain the shape of the graph during the following two time intervals: $t = 0$ to $t = 2.0$ min and then $t = 2.5$ and $t = 3.0$ min by referring to the volume of $\text{H}_2(\text{g})$ produced.

(IEB Paper 2, 2001)

3. Cobalt chloride crystals are dissolved in a beaker containing ethanol and then a few drops of water are added. After a period of time, the reaction reaches equilibrium as follows:



The solution, which is now just blue, is poured into three test tubes. State, in each case, what colour changes will be observed (if any) if the following are added in turn to each test tube:

- 1 cm^3 of distilled water
- A few crystals of sodium chloride
- The addition of dilute hydrochloric acid to the third test tube causes the solution to turn pink. Explain why this occurs.

(IEB Paper 2, 2001)

Chapter 4

Electrochemical Reactions - Grade 12

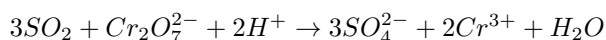
4.1 Introduction

Reaction Types in Grade 11 discussed *oxidation*, *reduction* and *redox reactions*. **Oxidation** involves a **loss of electrons** and **reduction** involves a **gain of electrons**. A **redox reaction** is a reaction where both oxidation and reduction take place. What is common to all of these processes is that they involve a *transfer of electrons* and a change in the oxidation state of the elements that are involved.



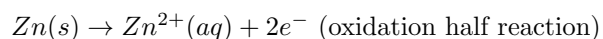
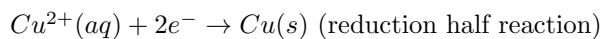
Exercise: Oxidation and reduction

1. Define the terms **oxidation** and **reduction**.
2. In each of the following reactions say whether the **iron** in the reactants is oxidised or reduced.
 - (a) $Fe \rightarrow Fe^{2+} + 2e^{-}$
 - (b) $Fe^{3+} + e^{-} \rightarrow Fe^{2+}$
 - (c) $Fe_2O_3 + 3CO \rightarrow 2Fe + 3CO_2$
 - (d) $Fe^{2+} \rightarrow Fe^{3+} + e^{-}$
 - (e) $Fe_2O_3 + 2Al \rightarrow Al_2O_3 + 2Fe$
3. In each of the following equations, say which elements in the reactants are oxidised and which are reduced.
 - (a) $CuO(s) + H_2(g) \rightarrow Cu(s) + H_2O(g)$
 - (b) $2NO(g) + 2CO(g) \rightarrow N_2(g) + 2CO_2(g)$
 - (c) $Mg(s) + FeSO_4(aq) \rightarrow MgSO_4(aq) + Fe(s)$
 - (d) $Zn(s) + 2AgNO_3(aq) \rightarrow 2Ag + Zn(NO_3)_2(aq)$
4. Which one of the substances listed below acts as the oxidising agent in the following reaction?

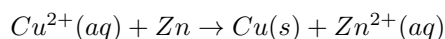


- (a) H^+
 - (b) Cr^{3+}
 - (c) SO_2
 - (d) $Cr_2O_7^{2-}$
-

In Grade 11, an experiment was carried out to see what happened when zinc granules are added to a solution of copper(II) sulphate. In the experiment, the Cu^{2+} ions from the copper(II) sulphate solution were reduced to copper metal, which was then deposited in a layer on the zinc granules. The zinc atoms were oxidised to form Zn^{2+} ions in the solution. The half reactions are as follows:



The overall redox reaction is:



There was an increase in the temperature of the reaction when you carried out this experiment. Is it possible that this heat energy could be converted into electrical energy? In other words, can we use a chemical reaction where there is an exchange of electrons, to produce electricity? And if this is possible, what would happen if an electrical current was supplied to *cause* some type of chemical reaction to take place?

An **electrochemical reaction** is a chemical reaction that *produces a voltage*, and therefore a flow of electrical current. An electrochemical reaction can also be the reverse of this process, in other words if an electrical current causes a chemical reaction to take place.



Definition: Electrochemical reaction

If a chemical reaction is caused by an external voltage, or if a voltage is caused by a chemical reaction, it is an electrochemical reaction.

Electrochemistry is the branch of chemistry that studies these electrochemical reactions. In this chapter, we will be looking more closely at different types of electrochemical reactions, and how these can be used in different ways.

4.2 The Galvanic Cell

Activity :: Experiment : Electrochemical reactions

Aim:

To investigate the reactions that take place in a zinc-copper cell

Apparatus:

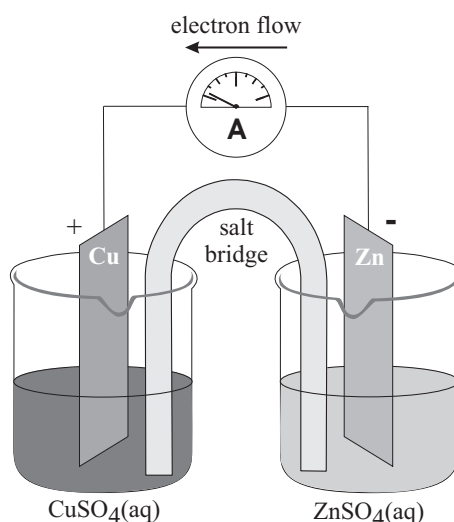
zinc plate, copper plate, measuring balance, zinc sulphate (ZnSO_4) solution (1 $\text{mol}\cdot\text{dm}^{-3}$), copper sulphate (CuSO_4) solution (1 $\text{mol}\cdot\text{dm}^{-3}$), two 250 ml beakers, U-tube, Na_2SO_4 solution, cotton wool, ammeter, connecting wire.

Method:

1. Measure the mass of the copper and zinc plates and record your findings.
2. Pour about 200 ml of the zinc sulphate solution into a beaker and put the zinc plate into it.
3. Pour about 200 ml of the copper sulphate solution into the second beaker and place the copper plate into it.
4. Fill the U-tube with the Na_2SO_4 solution and seal the ends of the tubes with the cotton wool. This will stop the solution from flowing out when the U-tube is turned upside down.

5. Connect the zinc and copper plates to the ammeter and observe whether the ammeter records a reading.
6. Place the U-tube so that one end is in the copper sulphate solution and the other end is in the zinc sulphate solution. Is there a reading on the ammeter? In which direction is the current flowing?
7. Take the ammeter away and connect the copper and zinc plates to each other directly using copper wire. Leave to stand for about one day.
8. After a day, remove the two plates and rinse them first with distilled water, then with alcohol and finally with ether. Dry the plates using a hair dryer.
9. Weigh the zinc and copper plates and record their mass. Has the mass of the plates changed from the original measurements?

Note: A voltmeter can also be used in place of the ammeter. A voltmeter will measure the potential difference across the cell.



Results:

During the experiment, you should have noticed the following:

- When the U-tube containing the Na_2SO_4 solution was absent, there was no reading on the ammeter.
- When the U-tube was connected, a reading was recorded on the ammeter.
- After the plates had been connected directly to each other and left for a day, there was a change in their mass. The mass of the zinc plate decreased, while the mass of the copper plate increased.
- The direction of electron flow is from the zinc plate towards the copper plate.

Conclusions:

When a zinc sulphate solution containing a zinc plate is connected by a U-tube to a copper sulphate solution containing a copper plate, reactions occur in both solutions. The decrease in mass of the zinc plate suggests that the zinc metal has been oxidised. The increase in mass of the copper plate suggests that reduction has occurred here to produce more copper metal. This will be explained in detail below.

4.2.1 Half-cell reactions in the Zn-Cu cell

The experiment above demonstrated a zinc-copper cell. This was made up of a zinc **half cell** and a copper **half cell**.

**Definition: Half cell**

A half cell is a structure that consists of a conductive electrode surrounded by a conductive electrolyte. For example, a zinc half cell could consist of a zinc metal plate (the electrode) in a zinc sulphate solution (the electrolyte).

How do we explain what has just been observed in the zinc-copper cell?

- *Copper plate*

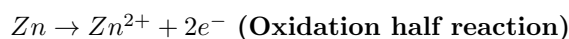
At the copper plate, there was an *increase* in mass. This means that Cu^{2+} ions from the copper sulphate solution were deposited onto the plate as atoms of copper metal. The half-reaction that takes place at the copper plate is:



Another shortened way to represent this copper half-cell is Cu^{2+}/Cu .

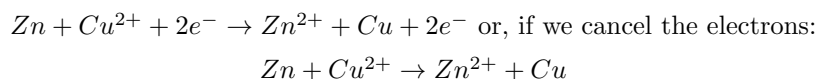
- *Zinc plate*

At the zinc plate, there was a *decrease* in mass. This means that some of the zinc goes into solution as Zn^{2+} ions. The electrons remain on the zinc plate, giving it a negative charge. The half-reaction that takes place at the zinc plate is:



The shortened way to represent the zinc half-cell is Zn/Zn^{2+} .

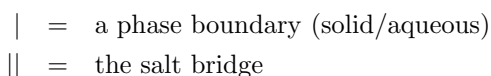
The overall reaction is:



For this electrochemical cell, the standard notation is:



where



In the notation used above, the oxidation half-reaction at the anode is written on the left, and the reduction half-reaction at the cathode is written on the right. In the Zn-Cu electrochemical cell, the direction of current flow in the external circuit is from the zinc electrode (where there has been a build up of electrons) to the copper electrode.

4.2.2 Components of the Zn-Cu cell

In the zinc-copper cell, the copper and zinc plates are called the **electrodes**. The electrode where oxidation occurs is called the **anode**, and the electrode where reduction takes place is called the **cathode**. In the zinc-copper cell, the zinc plate is the anode and the copper plate is the cathode.

**Definition: Electrode**

An electrode is an electrical conductor that is used to make contact with a metallic part of a circuit. The anode is the electrode where oxidation takes place. The cathode is the electrode where reduction takes place.

The zinc sulphate and copper sulphate solutions are called the **electrolyte** solutions.



Definition: Electrolyte

An electrolyte is a substance that contains free ions and which therefore behaves as an electrical conductor.

The U-tube also plays a very important role in the cell. In the Zn/Zn^{2+} half-cell, there is a build up of positive charge because of the release of electrons through oxidation. In the Cu^{2+}/Cu half-cell, there is a decrease in the positive charge because electrons are gained through reduction. This causes a movement of SO_4^{2-} ions into the beaker where there are too many positive ions, in order to neutralise the solution. Without this, the flow of electrons in the outer circuit stops completely. The U-tube is called the **salt bridge**. The salt bridge acts as a transfer medium that allows ions to flow through without allowing the different solutions to mix and react.



Definition: Salt bridge

A salt bridge, in electrochemistry, is a laboratory device that is used to connect the oxidation and reduction half-cells of a galvanic cell.

4.2.3 The Galvanic cell

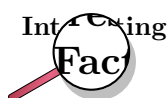
In the zinc-copper cell the important thing to notice is that the chemical reactions that take place at the two electrodes cause an electric current to flow through the outer circuit. In this type of cell, **chemical energy is converted to electrical energy**. These are called **galvanic cells**. The zinc-copper cell is one example of a galvanic cell. A galvanic cell (which is also sometimes referred to as a **voltaic** or **electrochemical** cell) consists of two metals that are connected by a salt bridge between the individual half-cells. A galvanic cell generates electricity using the reactions that take place at these two metals, each of which has a different **reaction potential**.

So what is meant by the 'reaction potential' of a substance? Every metal has a different half reaction and different dissolving rates. When two metals with different reaction potentials are used in a galvanic cell, a potential difference is set up between the two electrodes, and the result is a flow of current through the wire that connects the electrodes. In the zinc-copper cell, zinc has a higher reaction potential than copper and therefore dissolves more readily into solution. The metal 'dissolves' when it loses electrons to form positive metal ions. These electrons are then transferred through the connecting wire in the outer circuit.



Definition: Galvanic cell

A galvanic (voltaic) cell is an electrochemical cell that uses a chemical reaction between two dissimilar electrodes dipped in an electrolyte, to generate an electric current.

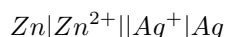


It was the Italian physician and anatomist Luigi Galvani who marked the birth of electrochemistry by making a link between chemical reactions and electricity. In 1780, Galvani discovered that when two different metals (copper and zinc for example) were connected together and then both touched to different parts of a nerve of a frog leg at the same time, they made the leg contract. He called this "animal electricity". While many scientists accepted his ideas, another scientist, Alessandro Volta, did not. In 1800, because of his professional disagreement over the galvanic response that had been suggested by Luigi Galvani, Volta developed the *voltaic pile*, which was very similar to the galvanic cell. It was the work of these two men that paved the way for all electrical batteries.



Worked Example 19: Understanding galvanic cells

Question: For the following cell:



1. Give the anode and cathode half-reactions.
2. Write the overall equation for the chemical reaction.
3. Give the direction of the current in the external circuit.

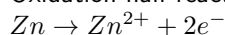
Answer

Step 1 : Identify the oxidation and reduction reactions

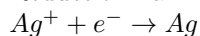
In the standard notation format, the oxidation reaction is written on the left and the reduction reaction on the right. So, in this cell, zinc is oxidised and silver ions are reduced.

Step 2 : Write the two half reactions

Oxidation half-reaction:

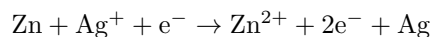


Reduction half-reaction:



Step 3 : Combine the half-reactions to get the overall equation.

When you combine the two half-reactions, all the reactants must go on the left side of the equation and the products must go on the right side of the equation. The overall equation therefore becomes:



Note that this equation is not balanced. This will be discussed later in the chapter.

Step 4 : Determine the direction of current flow

A build up of electrons occurs where oxidation takes place. This is at the zinc electrode. Current will therefore flow from the zinc electrode to the silver electrode.

4.2.4 Uses and applications of the galvanic cell

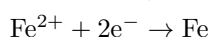
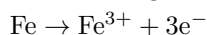
The principles of the galvanic cell are used to make **electrical batteries**. In science and technology, a battery is a device that stores chemical energy and makes it available in an electrical form. Batteries are made of electrochemical devices such as one or more galvanic cells, fuel cells or flow cells. Batteries have many uses including in torches, electrical appliances (long-life alkaline batteries), digital cameras (lithium batteries), hearing aids (silver-oxide batteries), digital watches (mercury batteries) and military applications (thermal batteries). Refer to chapter 5 for more information on batteries.

The galvanic cell can also be used for **electroplating**. Electroplating occurs when an electrically conductive object is coated with a layer of metal using electrical current. Sometimes, electroplating is used to give a metal particular properties such as corrosion protection or wear resistance. At other times, it can be for aesthetic reasons for example in the production of jewellery. This will be discussed in more detail later in this chapter.

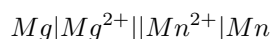


Exercise: Galvanic cells

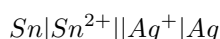
1. The following half-reactions take place in an electrochemical cell:



- Which is the oxidation half-reaction?
 - Which is the reduction half-reaction?
 - Name one oxidising agent.
 - Name one reducing agent.
 - Use standard notation to represent this electrochemical cell.
2. For the following cell:



- Give the cathode half-reaction.
 - Give the anode half-reaction.
 - Give the overall equation for the electrochemical cell.
 - What metals could be used for the electrodes in this electrochemical cell?
 - Suggest two electrolytes for this electrochemical cell.
 - In which direction will the current flow?
 - Draw a simple sketch of the complete cell.
3. For the following cell:



- Give the cathode half-reaction.
 - Give the anode half-reaction.
 - Give the overall equation for the electrochemical cell.
 - Draw a simple sketch of the complete cell.
-

4.3 The Electrolytic cell

In section 4.2, we saw that a chemical reaction that involves a transfer of electrons, can be used to produce an electric current. In this section, we are going to see whether the 'reverse' process applies. In other words, is it possible to use an electric current to force a particular chemical reaction to occur, which would otherwise not take place? The answer is 'yes', and the type of cell that is used to do this, is called an **electrolytic cell**.

Definition: Electrolytic cell

An electrolytic cell is a type of cell that uses electricity to drive a non-spontaneous reaction.

An electrolytic cell is activated by applying an electrical potential across the anode and cathode to force an internal chemical reaction between the ions that are in the electrolyte solution. This process is called **electrolysis**.

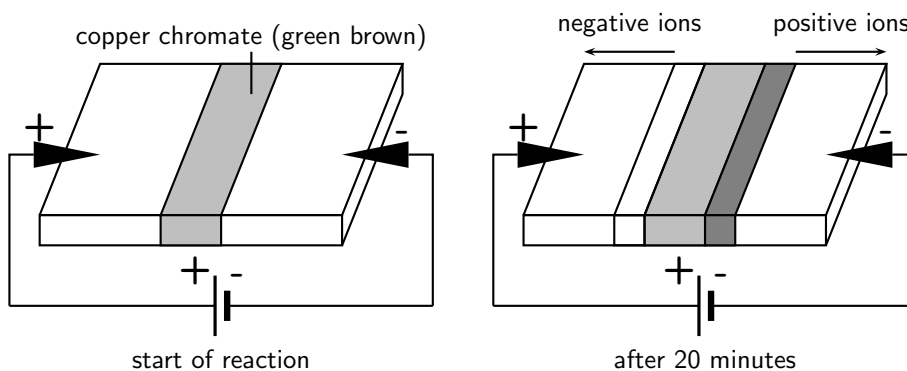


Definition: Electrolysis

In chemistry and manufacturing, electrolysis is a method of separating bonded elements and compounds by passing an electric current through them.

Activity :: Demonstration : The movement of coloured ions

A piece of filter paper is soaked in an ammonia-ammonium chloride solution and placed on a microscope slide. The filter paper is then connected to a supply of electric current using crocodile clips and connecting wire as shown in the diagram below. A line of copper chromate solution is placed in the centre of the filter paper. The colour of this solution is initially green-brown.



The current is then switched on and allowed to run for about 20 minutes. After this time, the central coloured band disappears and is replaced by two bands, one yellow and the other blue, which seem to have separated out from the first band of copper chromate.

Explanation:

- The cell that is used to supply an electric current sets up a potential difference across the circuit, so that one of the electrodes is positive and the other is negative.
- The chromate (CrO_4^{2-}) ions in the copper chromate solution are attracted to the positive electrode, while the Cu^{2+} ions are attracted to the negative electrode.

Conclusion:

The movement of ions occurs because the electric current in the outer circuit sets up a potential difference between the two electrodes.

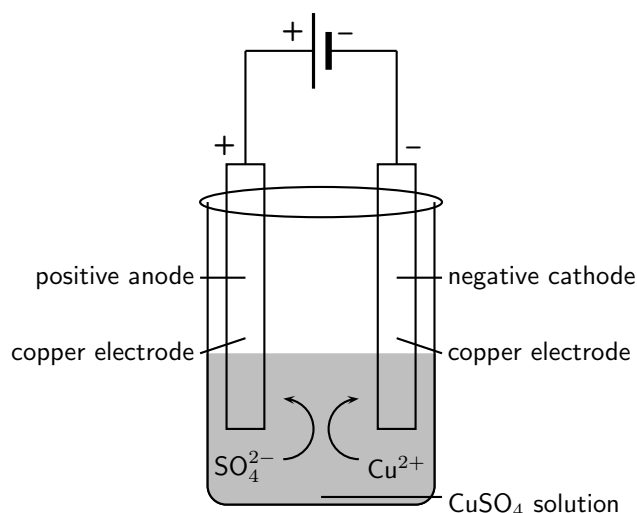
Similar principles apply in the electrolytic cell, where substances that are made of ions can be broken down into simpler substances through electrolysis.

4.3.1 The electrolysis of copper sulphate

There are a number of examples of electrolysis. The electrolysis of copper sulphate is just one.

Activity :: Demonstration : The electrolysis of copper sulphate

Two copper electrodes are placed in a solution of blue copper sulphate and are connected to a source of electrical current as shown in the diagram below. The current is turned on and the reaction is left for a period of time.

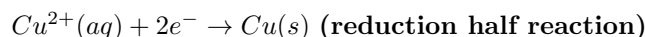


Observations:

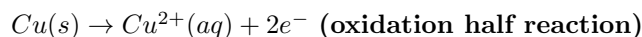
- The initial blue colour of the solution remains unchanged.
- It appears that copper has been *deposited* on one of the electrodes but *dissolved* from the other.

Explanation:

- At the negative cathode, positively charged Cu^{2+} ions are attracted to the negatively charged electrode. These ions gain electrons and are reduced to form copper metal, which is deposited on the electrode. The half-reaction that takes place is as follows:



- At the positive anode, copper metal is oxidised to form Cu^{2+} ions. This is why it appears that some of the copper has dissolved from the electrode. The half-reaction that takes place is as follows:



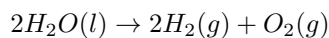
- The amount of copper that is *deposited* at one electrode is approximately the same as the amount of copper that is *dissolved* from the other. The number of Cu^{2+} ions in the solution therefore remains almost the same and the blue colour of the solution is unchanged.

Conclusion:

In this demonstration, an electric current was used to split CuSO_4 into its component ions, Cu^{2+} and SO_4^{2-} . This process is called *electrolysis*.

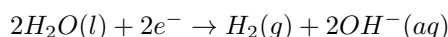
4.3.2 The electrolysis of water

Water can also undergo electrolysis to form hydrogen gas and oxygen gas according to the following reaction:

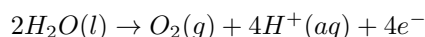


This reaction is very important because hydrogen gas has the potential to be used as an energy source. The electrolytic cell for this reaction consists of two electrodes (normally platinum metal), submerged in an electrolyte and connected to a source of electric current.

The reduction half-reaction that takes place at the cathode is as follows:



The oxidation half-reaction that takes place at the anode is as follows:



4.3.3 A comparison of galvanic and electrolytic cells

It should be much clearer now that there are a number of differences between a *galvanic* and an *electrolytic* cell. Some of these differences have been summarised in table 4.1.

Item	Galvanic cell	Electrolytic cell
Metals used for electrode	Two metals with different reaction potentials are used as electrodes	The same metal can be used for both the cathode and the anode
Charge of the anode	negative	positive
Charge of the cathode	positive	negative
The electrolyte solution/s	The electrolyte solutions are kept separate from one another, and are connected only by a salt bridge	The cathode and anode are in the same electrolyte
Energy changes	Chemical potential energy from chemical reactions is converted to electrical energy	An external supply of electrical energy causes a chemical reaction to occur
Applications	Run batteries, electroplating	Electrolysis e.g. of water, NaCl

Table 4.1: A comparison of galvanic and electrolytic cells



Exercise: Electrolysis

- An electrolytic cell consists of two electrodes in a silver chloride (AgCl) solution, connected to a source of current. A current is passed through the solution and Ag^+ ions are reduced to a silver metal deposit on one of the electrodes.
 - Give the equation for the reduction half-reaction.
 - Give the equation for the oxidation half-reaction.
- Electrolysis takes place in a solution of molten lead bromide (PbBr) to produce lead atoms.
 - Draw a simple diagram of the electrolytic cell.

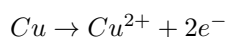
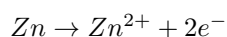
- (b) Give equations for the half-reactions that take place at the anode and cathode, and include these in the diagram.
- (c) On your diagram, show the direction in which current flows.
-

4.4 Standard Electrode Potentials

If a voltmeter is connected in the circuit of an electrochemical cell, a reading is obtained. In other words, there is a **potential difference** between the two half cells. In this section, we are going to look at this in more detail to try to understand more about the **electrode potentials** of each of the electrodes in the cell. We are going to break this section down so that you build up your understanding gradually. Make sure that you understand each subsection fully before moving on, otherwise it might get confusing!

4.4.1 The different reactivities of metals

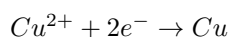
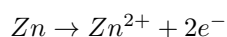
All metals have different reactivities. When metals react, they give away electrons and form positive ions. But some metals do this more easily than others. Look at the following two half reactions:



Of these two metals, zinc is more reactive and is more likely to give away electrons to form Zn^{2+} ions in solution, than is copper.

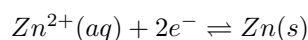
4.4.2 Equilibrium reactions in half cells

Let's think back to the Zn-Cu electrochemical cell. This cell is made up of two half cells and the reactions that take place at each of the electrodes are as follows:



At the zinc electrode, the zinc metal loses electrons and forms Zn^{2+} ions. The electrons are concentrated on the zinc metal while the Zn^{2+} ions are in solution. But some of the ions will be attracted back to the negatively charged metal, will gain their electrons again and will form zinc metal. A **dynamic equilibrium** is set up between the zinc metal and the Zn^{2+} ions in solution when the rate at which ions are *leaving* the metal is equal to the rate at which they are *joining* it again. The situation looks something like the diagram in figure 4.1.

The equilibrium reaction is represented like this:



(NOTE: By convention, the ions are written on the *left* hand side of the equation)

In the zinc half cell, the equilibrium lies far to the left because the zinc loses electrons easily to form Zn^{2+} ions. We can also say that the zinc is *oxidised* and that it is a strong *reducing agent*.

At the copper electrode, a similar process takes place. The difference though is that copper is not as reactive as zinc and so it does not form ions as easily. This means that the build up of

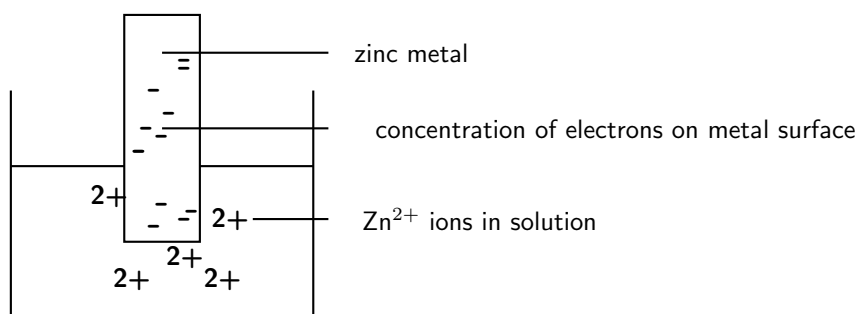


Figure 4.1: Zinc loses electrons to form positive ions in solution. The electrons accumulate on the metal surface.

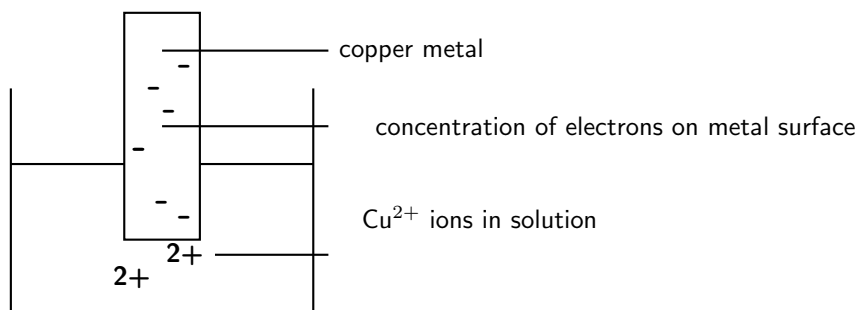
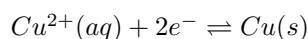


Figure 4.2: Zinc loses electrons to form positive ions in solution. The electrons accumulate on the metal surface.

electrons on the copper electrode is less (figure 4.2).

The equilibrium reaction is shown like this:



The equation lies far to the right because most of the copper is present as copper metal rather than as Cu^{2+} ions. In this half reaction, the Cu^{2+} ions are *reduced*.

4.4.3 Measuring electrode potential

If we put the two half cells together, a potential difference is set up in two places in the Zn-Cu cell:

1. There is a potential difference between the metal and the solution surrounding it because one is more negative than the other.
2. There is a potential difference between the Zn and Cu electrodes because one is more negative than the other.

It is the **potential difference** (recorded as a voltage) between the two electrodes that causes electrons, and therefore current, to flow from the more negative electrode to the less negative electrode.

The problem though is that we cannot measure the potential difference (voltage) between a metal and its surrounding solution in the cell. To do this, we would need to connect a voltmeter

to both the metal and the solution, which is not possible. This means we cannot measure the exact **electrode potential** (E°) of a particular metal. The electrode potential describes the ability of a metal to give up electrons. And if the exact electrode potential of each of the electrodes involved can't be measured, then it is difficult to calculate the potential difference between them. But what we *can* do is to try to describe the electrode potential of a metal *relative* to another substance. We need to use a **standard reference electrode** for this.

4.4.4 The standard hydrogen electrode

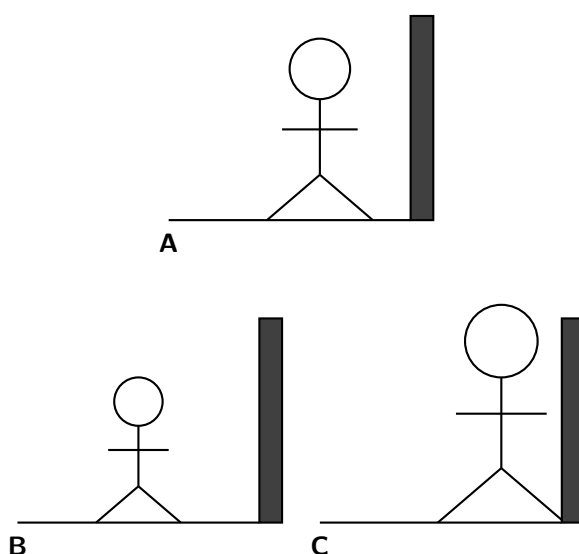
Before we look at the standard hydrogen electrode, it may be useful to have some more understanding of the ideas behind a 'reference electrode'. Refer to the Tip box on 'Understanding the ideas behind a reference electrode' before you read further.



Important: Understanding the ideas behind a reference electrode

Adapted from www.chemguide.co.uk

Let's say that you have a device that you can use to measure heights from some distance away. You want to use this to find out how tall a particular person is. Unfortunately, you can't see their feet because they are standing in long grass. Although you can't measure their absolute height, what you can do is to measure their height relative to the post next to them. Let's say that person A for example is 15 cm *shorter* than the height of the post. You could repeat this for a number of other people (B and C). Person B is 30 cm shorter than the post and person C is 10 cm *taller* than the post.



You could summarise your findings as follows:

Person	Height relative to post (cm)
A	-15
B	-30
C	+10

Although you don't know any of their absolute heights, you can rank them in order, and do some very simple sums to work out exactly how much taller one is than another. For example, person C is 25 cm taller than A and 40 cm taller than B.

As mentioned earlier, it is difficult to measure the *absolute* electrode potential of a particular substance, but we can use a reference electrode (similar to the 'post' in the Tip box example)

that we use to calculate *relative* electrode potentials for these substances. The reference electrode that is used is the **standard hydrogen electrode** (figure 4.3).

Definition: Standard hydrogen electrode

The standard hydrogen electrode is a redox electrode which forms the basis of the scale of oxidation-reduction potentials. The actual electrode potential of the hydrogen electrode is estimated to be 4.44 ± 0.02 V at 25°C , but its standard electrode potential is said to be zero at all temperatures so that it can be used as for comparison with other electrodes. The hydrogen electrode is based on the following redox half cell:

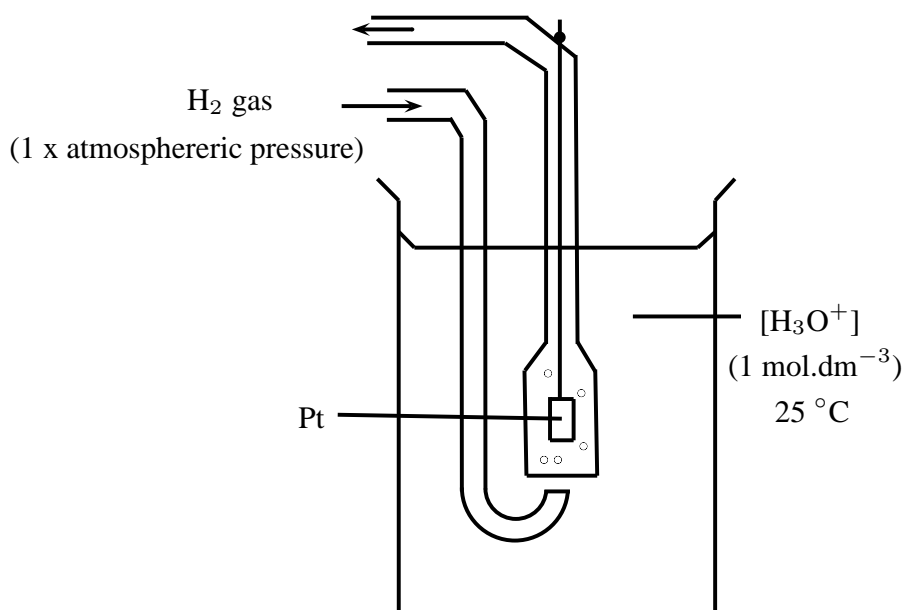
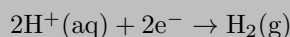
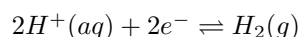


Figure 4.3: The standard hydrogen electrode

A standard hydrogen electrode consists of a platinum electrode in a solution containing H^+ ions. The solution (e.g. H_2SO_4) that contains the H^+ ions has a concentration of 1 mol.dm^{-3} . As the hydrogen gas bubbles over the platinum electrode, an equilibrium is set up between hydrogen molecules and hydrogen ions in solution. The reaction is as follows:

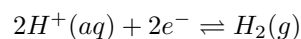
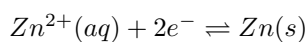


The position of this equilibrium can change if you change some of the conditions (e.g. concentration, temperature). It is therefore important that the conditions for the standard hydrogen electrode are standardised as follows: pressure = 100 kPa (1atm); temperature = 298 K (25°C) and concentration = 1 mol.dm^{-3} .

In order to *use* the hydrogen electrode, it needs to be attached to the electrode system that you are investigating. For example, if you are trying to determine the electrode potential of copper, you will need to connect the copper half cell to the hydrogen electrode; if you are trying to determine the electrode potential of zinc, you will need to connect the zinc half cell to the hydrogen electrode and so on. Let's look at the examples of zinc and copper in more detail.

1. Zinc

Zinc has a greater tendency than hydrogen to form ions, so if the standard hydrogen electrode is connected to the zinc half cell, the zinc will be relatively more negative because the electrons that are released when zinc is oxidised will accumulate on the metal. The equilibria on the two electrodes are as follows:



In the zinc half-reaction, the equilibrium lies far to the left and in the hydrogen half-reaction, the equilibrium lies far to the right. A simplified representation of the cell is shown in figure 4.4.

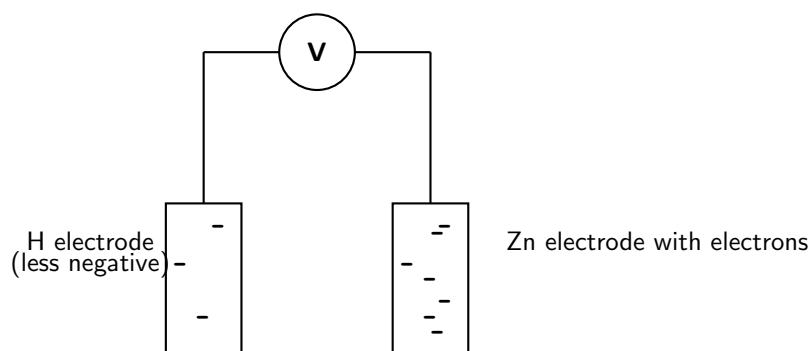
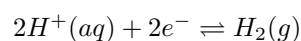
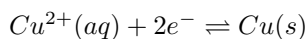


Figure 4.4: When zinc is connected to the standard hydrogen electrode, relatively few electrons build up on the platinum (hydrogen) electrode. There are lots of electrons on the zinc electrode.

The voltmeter measures the potential difference between the charge on these electrodes. In this case, the voltmeter would read 0.76 and would show that Zn is the negative electrode (i.e. it has a relatively higher number of electrons).

2. Copper

Copper has a lower tendency than hydrogen to form ions, so if the standard hydrogen electrode is connected to the copper half cell, the hydrogen will be relatively more negative. The equilibria on the two electrodes are as follows:



In the copper half-reaction, the equilibrium lies far to the right and in the hydrogen half-reaction, the equilibrium lies far to the left. A simplified representation of the cell is shown in figure 4.5.

The voltmeter measures the potential difference between the charge on these electrodes. In this case, the voltmeter would read 0.34 and would show that Cu is the positive electrode (i.e. it has a relatively lower number of electrons).

4.4.5 Standard electrode potentials

The voltages recorded earlier when zinc and copper were connected to a standard hydrogen electrode are in fact the **standard electrode potentials** for these two metals. It is important to remember that these are not *absolute* values, but are potentials that have been measured *relative* to the potential of hydrogen if the standard hydrogen electrode is taken to be zero.



Important: Conventions and voltage sign

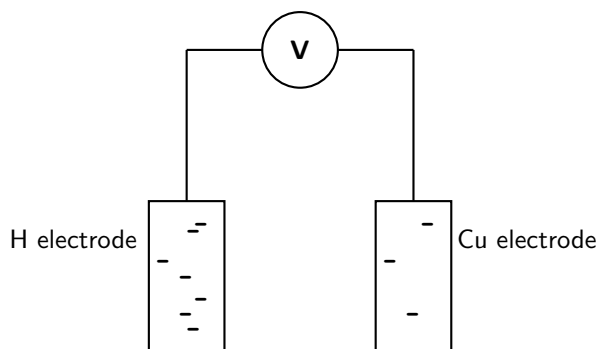


Figure 4.5: When copper is connected to the standard hydrogen electrode, relatively few electrons build up on the copper electrode. There are lots of electrons on the hydrogen electrode.

By convention, the hydrogen electrode is written on the *left hand side* of the cell. The sign of the voltage tells you the sign of the metal electrode.

In the examples we used earlier, zinc's electrode potential is actually -0.76 and copper is $+0.34$. So, if a metal has a *negative* standard electrode potential, it means it forms ions easily. The more negative the value, the easier it is for that metal to form ions. If a metal has a *positive* standard electrode potential, it means it does not form ions easily. This will be explained in more detail below.

Luckily for us, we do not have to calculate the standard electrode potential for every metal. This has been done already and the results are recorded in a table of standard electrode potentials (table 4.2).

A few examples from the table are shown in table 4.3. These will be used to explain some of the trends in the table of electrode potentials.

Refer to table 4.3 and notice the following trends:

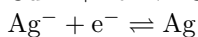
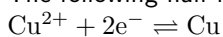
- Metals at the top of series (e.g. Li) have more negative values. This means they ionise easily, in other words, they release electrons easily. These metals are easily **oxidised** and are therefore good **reducing agents**.
- Metal ions at the bottom of the table are good at picking up electrons. They are easily **reduced** and are therefore good **oxidising agents**.
- The reducing ability (i.e. the ability to act as a reducing agent) of the metals in the table *increases* as you move *up* in the table.
- The oxidising ability of metals *increases* as you move *down* in the table.



Worked Example 20: Using the table of Standard Electrode Potentials

Question:

The following half-reactions take place in an electrochemical cell:



1. Which of these reactions will be the oxidation half-reaction in the cell?

Half-Reaction	E^0V
$Li^+ + e^- \rightleftharpoons Li$	-3.04
$K^+ + e^- \rightleftharpoons K$	-2.92
$Ba^{2+} + 2e^- \rightleftharpoons Ba$	-2.90
$Ca^{2+} + 2e^- \rightleftharpoons Ca$	-2.87
$Na^+ + e^- \rightleftharpoons Na$	-2.71
$Mg^{2+} + 2e^- \rightleftharpoons Mg$	-2.37
$Mn^{2+} + 2e^- \rightleftharpoons Mn$	-1.18
$2H_2O + 2e^- \rightleftharpoons H_2(g) + 2OH^-$	-0.83
$Zn^{2+} + 2e^- \rightleftharpoons Zn$	-0.76
$Cr^{2+} + 2e^- \rightleftharpoons Cr$	-0.74
$Fe^{2+} + 2e^- \rightleftharpoons Fe$	-0.44
$Cr^{3+} + 3e^- \rightleftharpoons Cr$	-0.41
$Cd^{2+} + 2e^- \rightleftharpoons Cd$	-0.40
$Co^{2+} + 2e^- \rightleftharpoons Co$	-0.28
$Ni^{2+} + 2e^- \rightleftharpoons Ni$	-0.25
$Sn^{2+} + 2e^- \rightleftharpoons Sn$	-0.14
$Pb^{2+} + 2e^- \rightleftharpoons Pb$	-0.13
$Fe^{3+} + 3e^- \rightleftharpoons Fe$	-0.04
$2H^+ + 2e^- \rightleftharpoons H_2(g)$	0.00
$S + 2H^+ + 2e^- \rightleftharpoons H_2S(g)$	0.14
$Sn^{4+} + 2e^- \rightleftharpoons Sn^{2+}$	0.15
$Cu^{2+} + e^- \rightleftharpoons Cu^+$	0.16
$SO_4^{2-} + 4H^+ + 2e^- \rightleftharpoons SO_2(g) + 2H_2O$	0.17
$Cu^{2+} + 2e^- \rightleftharpoons Cu$	0.34
$2H_2O + O_2 + 4e^- \rightleftharpoons 4OH^-$	0.40
$Cu^+ + e^- \rightleftharpoons Cu$	0.52
$I_2 + 2e^- \rightleftharpoons 2I^-$	0.54
$O_2(g) + 2H^+ + 2e^- \rightleftharpoons H_2O_2$	0.68
$Fe^{3+} + e^- \rightleftharpoons Fe^{2+}$	0.77
$NO_3^- + 2H^+ + e^- \rightleftharpoons NO_2(g) + H_2O$	0.78
$Hg^{2+} + 2e^- \rightleftharpoons Hg(l)$	0.78
$Ag^+ + e^- \rightleftharpoons Ag$	0.80
$NO_3^- + 4H^+ + 3e^- \rightleftharpoons NO(g) + 2H_2O$	0.96
$Br_2 + 2e^- \rightleftharpoons 2Br^-$	1.06
$O_2(g) + 4H^+ + 4e^- \rightleftharpoons 2H_2O$	1.23
$MnO_2 + 4H^+ + 2e^- \rightleftharpoons Mn^{2+} + 2H_2O$	1.28
$Cr_2O_7^{2-} + 14H^+ + 6e^- \rightleftharpoons 2Cr^{3+} + 7H_2O$	1.33
$Cl_2 + 2e^- \rightleftharpoons 2Cl^-$	1.36
$Au^{3+} + 3e^- \rightleftharpoons Au$	1.50
$MnO_4^- + 8H^+ + 5e^- \rightleftharpoons Mn^{2+} + 4H_2O$	1.52
$Co^{3+} + e^- \rightleftharpoons Co^{2+}$	1.82
$F_2 + 2e^- \rightleftharpoons 2F^-$	2.87

Table 4.2: Standard Electrode Potentials

Half-Reaction	E^0V
$Li^+ + e^- \rightleftharpoons Li$	-3.04
$Zn^{2+} + 2e^- \rightleftharpoons Zn$	-0.76
$Fe^{3+} + 3e^- \rightleftharpoons Fe$	-0.04
$2H^+ + 2e^- \rightleftharpoons H_2(g)$	0.00
$Cu^{2+} + 2e^- \rightleftharpoons Cu$	0.34
$Hg^{2+} + 2e^- \rightleftharpoons Hg(l)$	0.78
$Ag^+ + e^- \rightleftharpoons Ag$	0.80

Table 4.3: A few examples from the table of standard electrode potentials

2. Which of these reactions will be the reduction half-reaction in the cell?

Answer

Step 5 : Determine the electrode potential for each metal

From the table of standard electrode potentials, the electrode potential for the copper half-reaction is +0.34 V. The electrode potential for the silver half-reaction is +0.80 V.

Step 6 : Use the electrode potential values to determine which metal is oxidised and which is reduced

Both values are positive, but silver has a higher positive electrode potential than copper. This means that silver does not form ions easily, in other words, silver is more likely to be *reduced*. Copper is more likely to be *oxidised* and to form ions more easily than silver. Copper is the oxidation half-reaction and silver is the reduction half-reaction.

Important: Learning to understand the question in a problem.

Before you tackle this problem, make sure you understand exactly what the question is asking. If magnesium is able to displace silver from a solution of silver nitrate, this means that magnesium metal will form magnesium ions and the silver ions will become silver metal. In other words, there will now be *silver* metal and a solution of *magnesium nitrate*. This will only happen if magnesium has a greater tendency than silver to form ions. In other words, what the question is actually asking is whether magnesium or silver forms ions more easily.

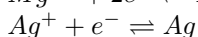
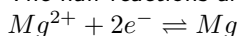
Worked Example 21: Using the table of Standard Electrode Potentials

Question: Is magnesium able to displace silver from a solution of silver nitrate?

Answer

Step 1 : Determine the half-reactions that would take place if magnesium were to displace silver nitrate.

The half-reactions are as follows:



Step 2 : Use the table of electrode potentials to see which metal forms ions more easily.

Looking at the electrode potentials for the magnesium and silver reactions:

For the magnesium half-reaction: $E^\circ V = -2.37$

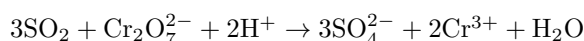
For the silver half-reaction: $E^\circ V = 0.80$

This means that magnesium is more easily **oxidised** than silver and the equilibrium in this half-reaction lies to the left. The oxidation reaction will occur spontaneously in magnesium. Silver is more easily **reduced** and the equilibrium lies to the right in this half-reaction. It can be concluded that magnesium will displace silver from a silver nitrate solution so that there is silver metal and magnesium ions in the solution.



Exercise: Table of Standard Electrode Potentials

- In your own words, explain what is meant by the 'electrode potential' of a metal.
- Give the standard electrode potential for each of the following metals:
 - magnesium
 - lead
 - nickel
- Refer to the electrode potentials in table 4.3.
 - Which of the metals is most likely to be oxidised?
 - Which metal is most likely to be reduced?
 - Which metal is the strongest reducing agent?
 - In the copper half-reaction, does the equilibrium position for the reaction lie to the left or to the right? Explain your answer.
 - In the mercury half-reaction, does the equilibrium position for the reaction lie to the left or to the right? Explain your answer.
 - If silver was added to a solution of copper sulphate, would it displace the copper from the copper sulphate solution? Explain your answer.
- Use the table of standard electrode potentials to put the following in order from the *strongest oxidising agent* to the *weakest oxidising agent*.
 - Cu^{2+}
 - MnO_4^-
 - Br_2
 - Zn^{2+}
- Look at the following half-reactions.
 - $\text{Ca}^{2+} + 2e^- \rightarrow \text{Ca}$
 - $\text{Cl}_2 + 2e^- \rightarrow 2\text{Cl}$
 - $\text{Fe}^{3+} + 3e^- \rightarrow \text{Fe}$
 - $\text{I}_2 + 2e^- \rightarrow 2\text{I}^-$
 - Which substance is the strongest oxidising agent?
 - Which substance is the strongest reducing agent?
- Which one of the substances listed below acts as the oxidising agent in the following reaction?

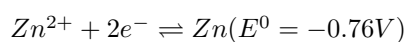


- (a) H^+
 - (b) Cr^{3+}
 - (c) SO_2
 - (d) $Cr_2O_7^{2-}$
- (IEB Paper 2, 2004)

7. If zinc is added to a solution of magnesium sulphate, will the zinc displace the magnesium from the solution? Give a detailed explanation for your answer.

4.4.6 Combining half cells

Let's stay with the example of the zinc and copper half cells. If we combine these cells as we did earlier in the chapter (section 4.2), the following two equilibria exist:



We know from demonstrations, and also by looking at the sign of the electrode potential, that when these two half cells are combined, zinc will be the oxidation half-reaction and copper will be the reduction half-reaction. A voltmeter connected to this cell will show that the zinc electrode is more negative than the copper electrode. The reading on the meter will show the potential difference between the two half cells. This is known as the **electromotive force (emf)** of the cell.



Definition: Electromotive Force (emf)

The emf of a cell is defined as the maximum potential difference between two electrodes or half cells in a voltaic cell. emf is the electrical driving force of the cell reaction. In other words, the higher the emf, the stronger the reaction.



Definition: Standard emf (E_{cell}^0)

Standard emf is the emf of a voltaic cell operating under standard conditions (i.e. 100 kPa, concentration = 1 mol.dm^{-3} and temperature = 298 K). The symbol 0 denotes standard conditions.

When we want to represent this cell, it is shown as follows:



The **anode** half cell (where *oxidation* takes place) is always written on the *left*. The **cathode** half cell (where *reduction* takes place) is always written on the *right*.

It is important to note that the potential difference across a cell is related to the extent to which the spontaneous cell reaction has reached equilibrium. In other words, as the reaction proceeds and the concentration of reactants decreases and the concentration of products increases, the reaction approaches equilibrium. When equilibrium is reached, the emf of the cell is zero and the cell is said to be 'flat'. There is no longer a potential difference between the two half cells, and therefore no more current will flow.

4.4.7 Uses of standard electrode potential

Standard electrode potentials have a number of different uses.

Calculating the emf of an electrochemical cell

To calculate the emf of a cell, you can use any one of the following equations:

$E_{(cell)}^0 = E^0 \text{ (right)} - E^0 \text{ (left)}$ ('right' refers to the electrode that is written on the right in standard cell notation. 'Left' refers to the half-reaction written on the left in this notation)

$E_{(cell)}^0 = E^0 \text{ (reduction half reaction)} - E^0 \text{ (oxidation half reaction)}$

$E_{(cell)}^0 = E^0 \text{ (oxidising agent)} - E^0 \text{ (reducing agent)}$

$E_{(cell)}^0 = E^0 \text{ (cathode)} - E^0 \text{ (anode)}$

So, for the Zn-Cu cell,

$$E_{(cell)}^0 = 0.34 - (-0.76)$$

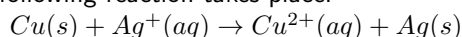
$$= 0.34 + 0.76$$

$$= 1.1 \text{ V}$$



Worked Example 22: Calculating the emf of a cell

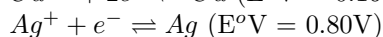
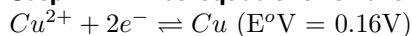
Question: The following reaction takes place:



1. Represent the cell using standard notation.
2. Calculate the cell potential (emf) of the electrochemical cell.

Answer

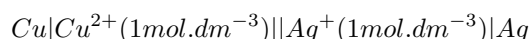
Step 1 : Write equations for the two half reactions involved



Step 2 : Determine which reaction takes place at the cathode and which is the anode reaction

Both half-reactions have positive electrode potentials, but the silver half-reaction has a higher positive value. In other words, silver does not form ions easily, and this must be the reduction half-reaction. Copper is the oxidation half-reaction. Copper is oxidised, therefore this is the anode reaction. Silver is reduced and so this is the cathode reaction.

Step 3 : Represent the cell using standard notation



Step 4 : Calculate the cell potential

$$E_{(cell)}^0 = E^0 \text{ (cathode)} - E^0 \text{ (anode)}$$

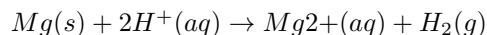
$$= +0.80 - (+0.34)$$

$$= +0.46 \text{ V}$$



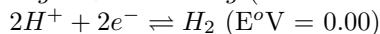
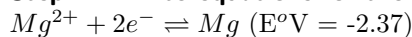
Worked Example 23: Calculating the emf of a cell

Question: Calculate the cell potential of the electrochemical cell in which the following reaction takes place, and represent the cell using standard notation.



Answer

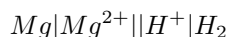
Step 1 : Write equations for the two half reactions involved



Step 2 : Determine which reaction takes place at the cathode and which is the anode reaction

From the overall equation, it is clear that magnesium is oxidised and hydrogen ions are reduced in this reaction. Magnesium is therefore the anode reaction and hydrogen is the cathode reaction.

Step 3 : Represent the cell using standard notation



Step 4 : Calculate the cell potential

$$E_{(cell)}^0 = E^0(\text{cathode}) - E^0(\text{anode})$$

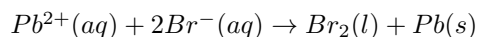
$$= 0.00 - (-2.37)$$

$$= +2.37 \text{ V}$$

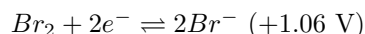
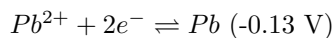
Predicting whether a reaction will take place spontaneously

Look at the following example to help you to understand how to predict whether a reaction will take place spontaneously or not.

In the reaction,



the two half reactions are as follows:



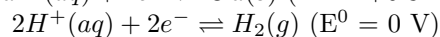
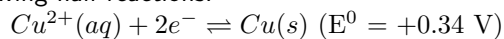
Important: Half cell reactions

You will see that the half reactions are written as they appear in the table of standard electrode potentials. It may be useful to highlight the reacting substance in each half reaction. In this case, the reactants are Pb^{2+} and Br^- ions.

Look at the electrode potential for the first half reaction. The negative value shows that lead loses electrons easily, in other words it is easily oxidised. The reaction would normally proceed from right to left (i.e. the equilibrium lies to the left), but in the original equation, the opposite is happening. It is the Pb^{2+} ions that are being reduced to lead. This part of the reaction is therefore not spontaneous. The positive electrode potential value for the bromine half-reaction shows that bromine is more easily reduced, in other words the equilibrium lies to the right. The spontaneous reaction proceeds from left to right. This is not what is happening in the original equation and therefore this is also not spontaneous. Overall it is clear then that the reaction will not proceed spontaneously.


Worked Example 24: Predicting whether a reaction is spontaneous

Question: Will copper react with dilute sulfuric acid (H_2SO_4)? You are given the following half reactions:


Answer

Step 5 : For each reaction, look at the electrode potentials and decide in which direction the equilibrium lies

In the first half reaction, the positive electrode potential means that copper does not lose electrons easily, in other words it is more easily reduced and the equilibrium position lies to the right. Another way of saying this is that the spontaneous reaction is the one that proceeds from left to right, when copper ions are reduced to copper metal.

In the second half reaction, the spontaneous reaction is from right to left.

Step 6 : Compare the equilibrium positions to the original reaction

What you should notice is that in the original reaction, the reactants are copper (Cu) and sulfuric acid (2H^+). During the reaction, the copper is oxidised and the hydrogen ions are reduced. But from an earlier step, we know that neither of these half reactions will proceed spontaneously in the direction indicated by the original reaction. The reaction is therefore not spontaneous.


Important:
A second method for predicting whether a reaction is spontaneous

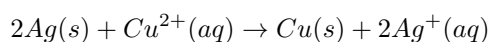
Another way of predicting whether a reaction occurs spontaneously, is to look at the sign of the emf value for the cell. If the emf is **positive** then the reaction is **spontaneous**. If the emf is **negative**, then the reaction is **not spontaneous**.

Balancing redox reactions

We will look at this in more detail in the next section.


Exercise: Predicting whether a reaction will take place spontaneously

1. Predict whether the following reaction will take place spontaneously or not. Show all your working.



2. Zinc metal reacts with an acid, H^+ (aq) to produce hydrogen gas.
 - (a) Write an equation for the reaction, using the table of electrode potentials.
 - (b) Predict whether the reaction will take place spontaneously. Show your working.
3. Four beakers are set up, each of which contains one of the following solutions:
 - (a) $\text{Mg}(\text{NO}_3)_2$

- (b) $\text{Ba}(\text{NO}_3)_2$
 (c) $\text{Cu}(\text{NO}_3)_2$
 (d) $\text{Al}(\text{NO}_3)_2$

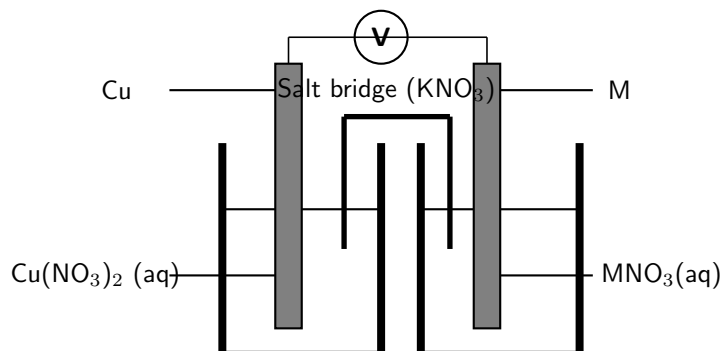
Iron is added to each of the beakers. In which beaker will a spontaneous reaction take place?

4. Which one of the following solutions can be stored in an aluminium container?
 (a) $\text{Cu}(\text{SO}_4)$
 (b) $\text{Zn}(\text{SO}_4)$
 (c) NaCl
 (d) $\text{Pb}(\text{NO}_3)_2$

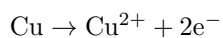


Exercise: Electrochemical cells and standard electrode potentials

1. An electrochemical cell is made up of a copper electrode in contact with a copper nitrate solution and an electrode made of an unknown metal M in contact with a solution of MNO_3 . A salt bridge containing a KNO_3 solution joins the two half cells. A voltmeter is connected across the electrodes. Under standard conditions the reading on the voltmeter is 0.46V.

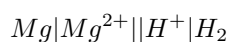


The reaction in the copper half cell is given by:

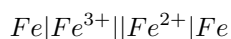


- (a) Write down the standard conditions which apply to this electrochemical cell.
 (b) Identify the metal M. Show calculations.
 (c) Use the standard electrode potentials to write down equations for the:
 i. cathode half-reaction
 ii. anode half-reaction
 iii. overall cell reaction
 (d) What is the purpose of the salt bridge?
 (e) Explain why a KCl solution would not be suitable for use in the salt bridge in this cell.
 (IEB Paper 2, 2004)
2. Calculate the emf for each of the following standard electrochemical cells:

- (a)



- (b)





3. Given the following two half-reactions:

- $Fe^{3+}(aq) + e^{-} \rightleftharpoons Fe^{2+}(aq)$
- $MnO_4^{-}(aq) + 8H^{+}(aq) + 5e^{-} \rightleftharpoons Mn^{2+}(aq) + 4H_2O(l)$

- (a) Give the standard electrode potential for each half-reaction.
- (b) Which reaction takes place at the cathode and which reaction takes place at the anode?
- (c) Represent the electrochemical cell using standard notation.
- (d) Calculate the emf of the cell

4.5 Balancing redox reactions

Half reactions can be used to balance redox reactions. We are going to use some worked examples to help explain the method.

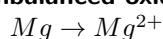


Worked Example 25: Balancing redox reactions

Question: Magnesium reduces copper (II) oxide to copper. In the process, magnesium is oxidised to magnesium ions. Write a balanced equation for this reaction.

Answer

Step 1 : Write down the unbalanced oxidation half reaction.

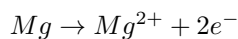


Step 2 : Balance the number of atoms on both sides of the equation.

You are allowed to add hydrogen ions (H^{+}) and water molecules if the reaction takes place in an acid medium. If the reaction takes place in a basic medium, you can add either hydroxide ions (OH^{-}) or water molecules. In this case, there is one magnesium atom on the left and one on the right, so no additional atoms need to be added.

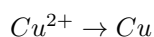
Step 3 : Once the atoms are balanced, check that the charges balance.

Charges can be balanced by adding electrons to either side. The charge on the left of the equation is 0, but the charge on the right is +2. Therefore, two electrons must be added to the right hand side so that the charges balance. The half reaction is now:

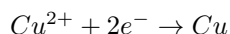


Step 4 : Repeat the above steps, but this time using the reduction half reaction.

The reduction half reaction is:



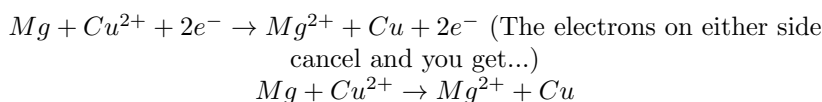
The atoms balance but the charges don't. Two electrons must be added to the right hand side.



Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.

No multiplication is needed because there are two electrons on either side.

Step 6 : Combine the two half reactions to get a final equation for the overall reaction.



Step 7 : Do a final check to make sure that the equation is balanced
In this case, it is.

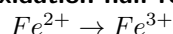


Worked Example 26: Balancing redox reactions

Question: Chlorine gas oxidises Fe(II) ions to Fe(III) ions. In the process, chlorine is reduced to chloride ions. Write a balanced equation for this reaction.

Answer

Step 1 : Write down the oxidation half reaction.

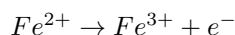


Step 2 : Balance the number of atoms on both sides of the equation.

There is one iron atom on the left and one on the right, so no additional atoms need to be added.

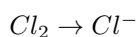
Step 3 : Once the atoms are balanced, check that the charges balance.

The charge on the left of the equation is +2, but the charge on the right is +3. Therefore, one electron must be added to the right hand side so that the charges balance. The half reaction is now:

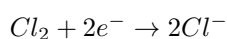


Step 4 : Repeat the above steps, but this time using the reduction half reaction.

The reduction half reaction is:

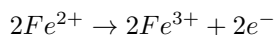


The atoms don't balance, so we need to multiply the right hand side by two to fix this. Two electrons must be added to the left hand side to balance the charges.

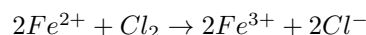


Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.

We need to multiply the oxidation half reaction by two so that the number of electrons on either side are balanced. This gives:



Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

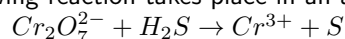


Step 7 : Do a final check to make sure that the equation is balanced
The equation is balanced.



Worked Example 27: Balancing redox reactions in an acid medium

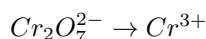
Question: The following reaction takes place in an acid medium:



Write a balanced equation for this reaction.

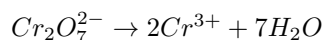
Answer

Step 1 : Write down the oxidation half reaction.

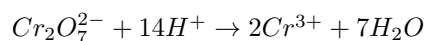


Step 2 : Balance the number of atoms on both sides of the equation.

We need to multiply the right side by two so that the number of Cr atoms will balance. To balance the oxygen atoms, we will need to add water molecules to the right hand side.

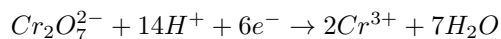


Now the oxygen atoms balance but the hydrogens don't. Because the reaction takes place in an acid medium, we can add hydrogen ions to the left side.



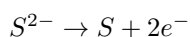
Step 3 : Once the atoms are balanced, check that the charges balance.

The charge on the left of the equation is $(-2+14) = +12$, but the charge on the right is $+6$. Therefore, six electrons must be added to the left hand side so that the charges balance. The half reaction is now:



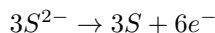
Step 4 : Repeat the above steps, but this time using the reduction half reaction.

The reduction half reaction after the charges have been balanced is:

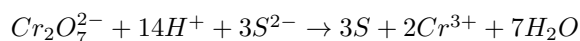


Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.

We need to multiply the reduction half reaction by three so that the number of electrons on either side are balanced. This gives:



Step 6 : Combine the two half reactions to get a final equation for the overall reaction.



Step 7 : Do a final check to make sure that the equation is balanced

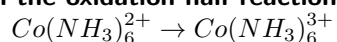


Worked Example 28: Balancing redox reactions in an alkaline medium

Question: If ammonia solution is added to a solution that contains cobalt(II) ions, a complex ion is formed, called the hexaamminecobalt(II) ion ($Co(NH_3)_6^{2+}$). In a chemical reaction with hydrogen peroxide solution, hexaamminecobalt ions are oxidised by hydrogen peroxide solution to the hexaamminecobalt(III) ion $Co(NH_3)_6^{3+}$. Write a balanced equation for this reaction.

Answer

Step 1 : Write down the oxidation half reaction

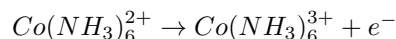


Step 2 : Balance the number of atoms on both sides of the equation.

The number of atoms are the same on both sides.

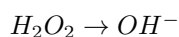
Step 3 : Once the atoms are balanced, check that the charges balance.

The charge on the left of the equation is +2, but the charge on the right is +3. One electron must be added to the right hand side to balance the charges in the equation. The half reaction is now:

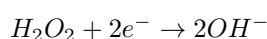


Step 4 : Repeat the above steps, but this time using the reduction half reaction.

Although you don't actually know what product is formed when hydrogen peroxide is reduced, the most logical product is OH^{-} . The reduction half reaction is:

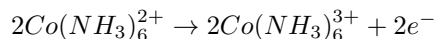


After the atoms and charges have been balanced, the final equation for the reduction half reaction is:

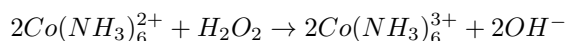


Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.

We need to multiply the oxidation half reaction by two so that the number of electrons on both sides are balanced. This gives:



Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

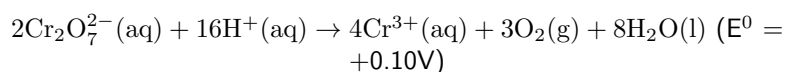


Step 7 : Do a final check to make sure that the equation is balanced



Exercise: Balancing redox reactions

- Balance the following equations.
 - $HNO_3 + PbS \rightarrow PbSO_4 + NO + H_2O$
 - $NaI + Fe_2(SO_4)_3 \rightarrow I_2 + FeSO_4 + Na_2SO_4$
- Manganate(VII) ions (MnO_4^-) oxidise hydrogen peroxide (H_2O_2) to oxygen gas. The reaction is done in an acid medium. During the reaction, the manganate(VII) ions are reduced to manganese(II) ions (Mn^{2+}). Write a balanced equation for the reaction.
- Chlorine gas is prepared in the laboratory by adding concentrated hydrochloric acid to manganese dioxide powder. The mixture is carefully heated.
 - Write down a balanced equation for the reaction which takes place.
 - Using standard electrode potentials, show by calculations why this mixture needs to be heated.
 - Besides chlorine gas which is formed during the reaction, hydrogen chloride gas is given off when the concentrated hydrochloric acid is heated. Explain why the hydrogen chloride gas is removed from the gas mixture when the gas is bubbled through water.
(IEB Paper 2, 2004)
- The following equation can be deduced from the table of standard electrode potentials:



This equation implies that an acidified solution of aqueous potassium dichromate (orange) should react to form Cr^{3+} (green). Yet aqueous laboratory solutions of potassium dichromate remain orange for years. Which ONE of the following best explains this?

- Laboratory solutions of aqueous potassium dichromate are not acidified
 - The E^0 value for this reaction is only +0.10V
 - The activation energy is too low
 - The reaction is non-spontaneous
- (IEB Paper 2, 2002)
- Sulfur dioxide gas can be prepared in the laboratory by heating a mixture of copper turnings and concentrated sulfuric acid in a suitable flask.

- Derive a balanced ionic equation for this reaction using the half-reactions that take place.
- Give the E^0 value for the overall reaction.
- Explain why it is necessary to heat the reaction mixture.
- The sulfur dioxide gas is now bubbled through an aqueous solution of potassium dichromate. Describe and explain what changes occur during this process.

(IEB Paper 2, 2002)

4.6 Applications of electrochemistry

Electrochemistry has a number of different uses, particularly in industry. We are going to look at a few examples.

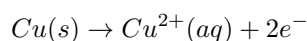
4.6.1 Electroplating

Electroplating is the process of using electrical current to coat an electrically conductive object with a thin layer of metal. Mostly, this application is used to deposit a layer of metal that has some desired property (e.g. abrasion and wear resistance, corrosion protection, improvement of aesthetic qualities etc.) onto a surface that doesn't have that property. Electro-refining (also sometimes called *electrowinning* is electroplating on a large scale. Electrochemical reactions are used to deposit pure metals from their ores. One example is the electrorefining of copper.

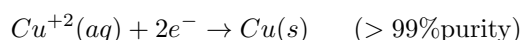
Copper plays a major role in the electrical reticulation industry as it is very conductive and is used in electric cables. One of the problems though is that copper must be pure if it is to be an effective current carrier. One of the methods used to purify copper, is electrowinning. The copper electrowinning process is as follows:

- Bars of crude (impure) copper containing other metallic impurities is placed on the *anodes*.
- The *cathodes* are made up of pure copper with few impurities.
- The electrolyte is a solution of aqueous $CuSO_4$ and H_2SO_4 .
- When current passes through the cell, electrolysis takes place. The impure copper anode dissolves to form Cu^{2+} ions in solution. These positive ions are attracted to the negative cathode, where reduction takes place to produce pure copper metal. The reactions that take place are as follows:

At the *anode*:



At the *cathode*:



- The other metal impurities (Zn, Au, Ag, Fe and Pb) do not dissolve and form a solid sludge at the bottom of the tank or remain in solution in the electrolyte.

4.6.2 The production of chlorine

Electrolysis can also be used to produce chlorine gas from brine/seawater (NaCl). This is sometimes referred to as the 'Chlor-alkali' process. The reactions that take place are as follows:

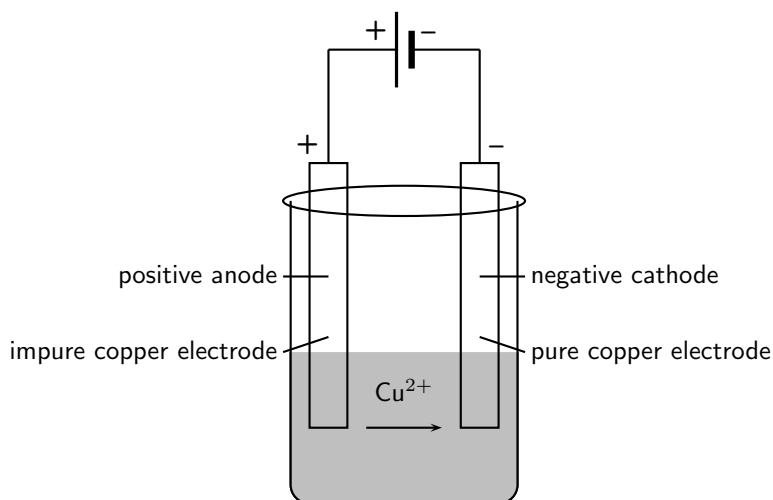
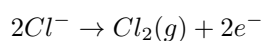
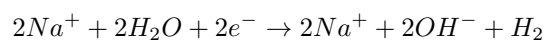


Figure 4.6: A simplified diagram to illustrate what happens during the electrowinning of copper

At the *anode* the reaction is:



whereas at the *cathode*, the following happens:



The overall reaction is:

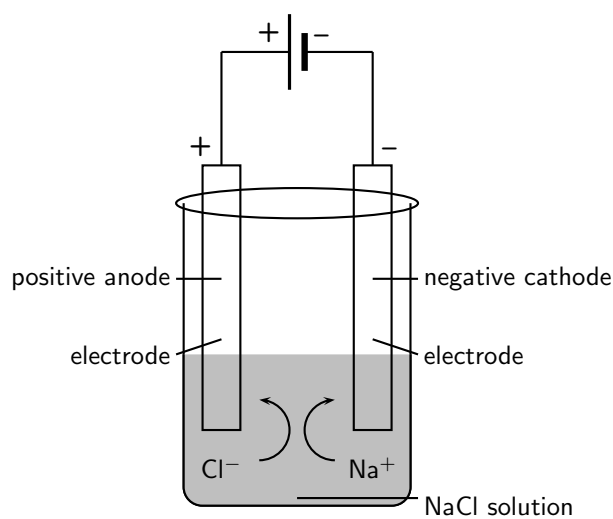
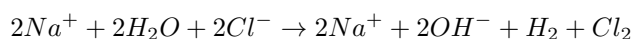


Figure 4.7: The electrolysis of sodium chloride

Chlorine is a very important chemical. It is used as a bleaching agent, a disinfectant, in solvents, pharmaceuticals, dyes and even plastics such as polyvinylchloride (PVC).

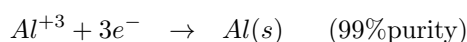
4.6.3 Extraction of aluminium

Aluminum metal is a commonly used metal in industry where its properties of being both light and strong can be utilized. It is also used in the manufacture of products such as aeroplanes and motor cars. The metal is present in deposits of bauxite which is a mixture of silicas, iron oxides and hydrated alumina ($Al_2O_3 \times H_2O$).

Electrolysis can be used to extract aluminum from bauxite. The process described below produces 99% pure aluminum:

1. Aluminum is melted along with cryolite (Na_3AlF_6) which acts as the electrolyte. Cryolite helps to lower the melting point and dissolve the ore.
2. The *anode* carbon rods provide sites for the oxidation of O^{2-} and F^- ions. Oxygen and fluorine gas are given off at the anodes and also lead to anode consumption.
3. At the *cathode* cell lining, the Al^{3+} ions are reduced and metal aluminum deposits on the lining.
4. The AlF_6^{3-} electrolyte is stable and remains in its molten state.

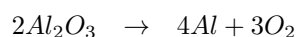
The basic electrolytic reactions involved are as follows: At the *cathode*:



At the *anode*:



The overall reaction is as follows:



The only problem with this process is that the reaction is *endothermic* and large amounts of electricity are needed to drive the reaction. The process is therefore very expensive.

4.7 Summary

- An **electrochemical reaction** is one where either a chemical reaction produces an external voltage, or where an external voltage causes a chemical reaction to take place.
- In a **galvanic cell** a chemical reaction produces a current in the external circuit. An example is the zinc-copper cell.
- A galvanic cell has a number of **components**. It consists of two **electrodes**, each of which is placed in a separate beaker in an **electrolyte** solution. The two electrolytes are connected by a **salt bridge**. The electrodes are connected to each other by an external circuit wire.
- One of the electrodes is the **anode**, where **oxidation** takes place. The **cathode** is the electrode where **reduction** takes place.
- In a galvanic cell, the build up of electrons at the anode sets up a potential difference between the two electrodes, and this causes a **current** to flow in the external circuit.
- A **galvanic cell** is therefore an electrochemical cell that uses a chemical reaction between two dissimilar electrodes dipped in an electrolyte to generate an electric current.
- The standard notation for a galvanic cell such as the zinc-copper cell is as follows:



where

| = a phase boundary (solid/aqueous)

|| = the salt bridge

- The galvanic cell is used in **batteries** and in **electroplating**.
- An **electrolytic cell** is an electrochemical cell that uses electricity to drive a non-spontaneous reaction. In an electrolytic cell, **electrolysis** occurs, which is a process of separating elements and compounds using an electric current.
- One example of an electrolytic cell is the electrolysis of copper sulphate to produce copper and sulphate ions.
- Different metals have different **reaction potentials**. The reaction potential of metals (in other words, their ability to ionise), is recorded in a **standard table of electrode potential**. The more negative the value, the greater the tendency of the metal to be oxidised. The more positive the value, the greater the tendency of the metal to be reduced.
- The values on the standard table of electrode potentials are measured relative to the **standard hydrogen electrode**.
- The emf of a cell can be calculated using one of the following equations:

$$E_{(cell)}^0 = E^0(\text{right}) - E^0(\text{left})$$

$$E_{(cell)}^0 = E^0(\text{reduction half reaction}) - E^0(\text{oxidation half reaction})$$

$$E_{(cell)}^0 = E^0(\text{oxidising agent}) - E^0(\text{reducing agent})$$

$$E_{(cell)}^0 = E^0(\text{cathode}) - E^0(\text{anode})$$
- It is possible to predict whether a reaction is **spontaneous** or not, either by looking at the sign of the cell's emf or by comparing the electrode potentials of the two half cells.
- It is possible to **balance** redox equations using the half-reactions that take place.
- There are a number of important **applications** of electrochemistry. These include **electroplating**, the **production of chlorine** and the **extraction of aluminium**.



Exercise: Summary exercise

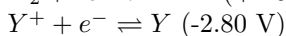
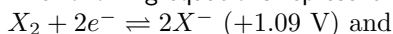
- For each of the following, say whether the statement is **true** or **false**. If it is false, re-write the statement correctly.
 - The anode in an electrolytic cell has a negative charge.
 - The reaction $2\text{KClO}_3 \rightarrow 2\text{KCl} + 3\text{O}_2$ is an example of a redox reaction.
 - Lead is a stronger oxidising agent than nickel.
- For each of the following questions, choose the **one correct** answer.
 - Which one of the following reactions is a redox reaction?
 - $\text{HCl} + \text{NaOH} \rightarrow \text{NaCl} + \text{H}_2\text{O}$
 - $\text{AgNO}_3 + \text{NaI} \rightarrow \text{AgI} + \text{NaNO}_3$
 - $2\text{FeCl}_3 + 2\text{H}_2\text{O} + \text{SO}_2 \rightarrow \text{H}_2\text{SO}_4 + 2\text{HCl} + 2\text{FeCl}_2$
 - $\text{BaCl}_2 + \text{MgSO}_4 \rightarrow \text{MgCl}_2 + \text{BaSO}_4$
 (IEB Paper 2, 2003)
 - Consider the reaction represented by the following equation:

$$\text{Br}_{2(l)} + 2\text{I}_{aq}^- \rightarrow 2\text{Br}_{aq}^- + \text{I}_{2(s)}$$
 Which one of the following statements about this reaction is correct?
 - bromine is oxidised
 - bromine acts as a reducing agent

- iii. the iodide ions are oxidised
- iv. iodine acts as a reducing agent

(IEB Paper 2, 2002)

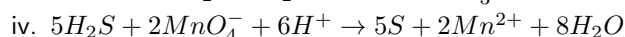
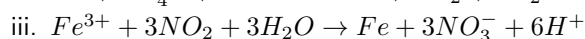
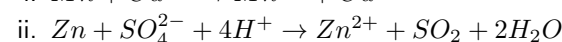
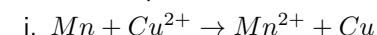
- (c) The following equations represent two hypothetical half-reactions:



Which one of the following substances from these half-reactions has the greatest tendency to donate electrons?

- i. X^-
- ii. X_2
- iii. Y
- iv. Y^+

- (d) Which one of the following redox reactions will **not** occur spontaneously at room temperature?



- (e) Which statement is CORRECT for a Zn-Cu galvanic cell that operates under standard conditions?

- i. The concentration of the Zn^{2+} ions in the zinc half-cell gradually decreases.
- ii. The concentration of the Cu^{2+} ions in the copper half-cell gradually increases.
- iii. Negative ions migrate from the zinc half-cell to the copper half-cell.
- iv. The intensity of the colour of the electrolyte in the copper half-cell gradually decreases.

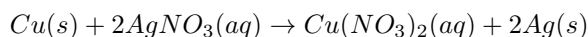
(DoE Exemplar Paper 2, 2008)

3. In order to investigate the rate at which a reaction proceeds, a learner places a beaker containing concentrated nitric acid on a sensitive balance. A few pieces of copper metal are dropped into the nitric acid.

- (a) Use the relevant half-reactions from the table of Standard Reduction Potentials to derive the balanced nett ionic equation for the reaction that takes place in the beaker.
- (b) What chemical property of nitric acid is illustrated by this reaction?
- (c) List three observations that this learner would make during the investigation.

(IEB Paper 2, 2005)

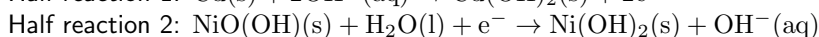
4. The following reaction takes place in an electrochemical cell:



- (a) Give an equation for the oxidation half reaction.
- (b) Which metal is used as the anode?
- (c) Determine the emf of the cell under standard conditions.

(IEB Paper 2, 2003)

5. The nickel-cadmium (NiCad) battery is small and light and is made in a sealed unit. It is used in portable appliances such as calculators and electric razors. The following two half reactions occur when electrical energy is produced by the cell.

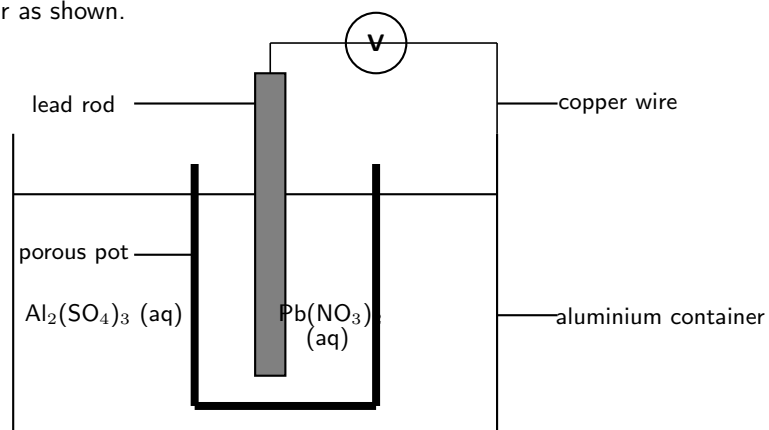


- (a) Which half reaction (1 or 2) occurs at the anode? Give a reason for your answer.
- (b) Which substance is oxidised?

- (c) Derive a balanced ionic equation for the overall cell reaction for the discharging process.
- (d) Use your result above to state in which direction the cell reaction will proceed (forward or reverse) when the cell is being charged.

(IEB Paper 2, 2001)

6. An electrochemical cell is constructed by placing a lead rod in a porous pot containing a solution of lead nitrate (see sketch). The porous pot is then placed in a large aluminium container filled with a solution of aluminium sulphate. The lead rod is then connected to the aluminium container by a copper wire and voltmeter as shown.



- (a) Define the term *reduction*.
- (b) In which direction do electrons flow in the copper wire? (Al to Pb or Pb to Al)
- (c) Write balanced equations for the reactions that take place at...
- the anode
 - the cathode
- (d) Write a balanced net ionic equation for the reaction which takes place in this cell.
- (e) What are the two functions of the porous pot?
- (f) Calculate the emf of this cell under standard conditions.

(IEB Paper 2, 2005)

Chapter 5

The Chemical Industry - Grade 12

5.1 Introduction

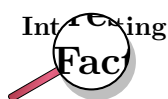
The chemical industry has been around for a very long time, but not always in the way we think of it today! Dyes, perfumes, medicines and soaps are all examples of products that have been made from chemicals that are found in either plants or animals. However, it was not until the time of the Industrial Revolution that the chemical industry as we know it today began to develop. At the time of the Industrial Revolution, the human population began to grow very quickly and more and more people moved into the cities to live. With this came an increase in the need for things like paper, glass, textiles and soaps. On the farms, there was a greater demand for fertilisers to help produce enough food to feed all the people in cities and rural areas. Chemists and engineers responded to these growing needs by using their technology to produce a variety of new chemicals. This was the start of the chemical industry.

In South Africa, the key event that led to the growth of the chemical industry was the discovery of diamonds and gold in the late 1800's. Mines needed explosives so that they could reach the diamonds and gold-bearing rock, and many of the main chemical companies in South Africa developed to meet this need for explosives. In this chapter, we are going to take a closer look at one of South Africa's major chemical companies, **Sasol**, and will also explore the **chloralkali** and **fertiliser** industries.

5.2 Sasol

Oil and natural gas are important fuel resources. Unfortunately, South Africa has no large oil reserves and, until recently, had very little natural gas. One thing South Africa *does* have however, is large supplies of **coal**. Much of South Africa's chemical industry has developed because of the need to produce oil and gas from coal, and this is where Sasol has played a very important role.

Sasol was established in 1950, with its main aim being to convert low grade coal into petroleum (crude oil) products and other chemical feedstocks. A 'feedstock' is something that is used to make another product. Sasol began producing oil from coal in 1955.



The first interest in coal chemistry started as early as the 1920's. In the early 1930's a research engineer called Etienne Rousseau was employed to see whether oil could be made from coal using a new German technology called the **Fischer-Tropsch** process. After a long time, and after many negotiations, Rousseau was given the rights to operate a plant using this new process. As a result, the government-sponsored 'South African Coal, Oil and Gas Corporation Ltd'

(commonly called 'Sasol') was formed in 1950 to begin making oil from coal. A manufacturing plant was established in the Free State and the town of **Sasolburg** developed around this plant. Production began in 1955. In 1969, the **Natref** crude oil refinery was established, and by 1980 and 1982 Sasol Two and Sasol Three had been built at **Secunda**.

5.2.1 Sasol today: Technology and production

Today, Sasol is an oil and gas company with diverse chemical interests. Sasol has three main areas of operation: Firstly, **coal to liquid fuels technology**, secondly the production of **crude oil** and thirdly the conversion of **natural gas to liquid fuel**.

1. Coal to liquid fuels

Sasol is involved in mining coal and converting it into synthetic fuels, using the **Fischer-Tropsch** technology. Figure 5.1 is a simplified diagram of the process that is involved.

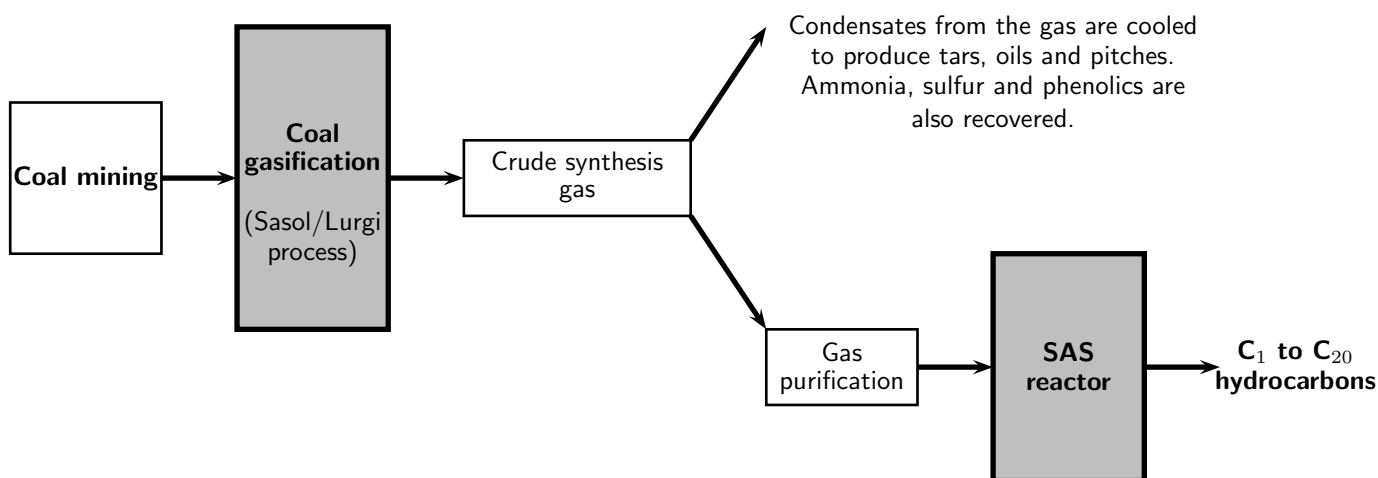
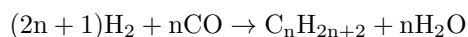


Figure 5.1: The gasification of coal to produce liquid fuels

Coal gasification is also known as the **Sasol/Lurgi** gasification process, and involves converting low grade coal to a synthesis gas. Low grade coal has a low percentage carbon, and contains other impurities. The coal is put under extremely high pressure and temperature in the presence of steam and oxygen. The gas that is produced has a high concentration of hydrogen (H_2) and carbon monoxide (CO). That is why it is called a 'synthesis gas', because it is a mixture of more than one gas.

In the **Sasol Advanced Synthol (SAS) reactors**, the gas undergoes a high temperature Fischer-Tropsch conversion. Hydrogen and carbon monoxide react under high pressure and temperature and in the presence of an iron catalyst, to produce a range of hydrocarbon products. Below is the generalised equation for the process. Don't worry too much about the numbers that you see in front of the reactants and products. It is enough just to see that the reaction of hydrogen and carbon monoxide (the two gases in the *synthesis gas*) produces a hydrocarbon and water.



A range of hydrocarbons are produced, including petrol, diesel, jet fuel, propane, butane, ethylene, polypropylene, alcohols and acetic acids.



Important: Different types of fuels

It is important to understand the difference between types of fuels and the terminology that is used for them. The table below summarises some of the fuels that will be mentioned in this chapter.

Compound	Description
Petroleum (crude oil)	A naturally occurring liquid that forms in the earth's lithosphere (see Grade 11 notes). It is a mixture of hydrocarbons, mostly alkanes, ranging from C_5H_{12} to $C_{18}H_{38}$.
Natural gas	Natural gas has the same origin as petroleum, but is made up of shorter hydrocarbon chains.
Paraffin wax	This is made up of longer hydrocarbon chains, making it a solid compound.
Petrol (gasoline)	A liquid fuel that is derived from petroleum, but which contains extra additives to increase the octane rating of the fuel. Petrol is used as a fuel in combustion engines.
Diesel	Diesel is also derived from petroleum, but is used in diesel engines.
Liquid Petroleum Gas (LPG)	LPG is a mixture of hydrocarbon gases, and is used as a fuel in heating appliances and vehicles. Some LPG mixtures contain mostly propane, while others are mostly butane. LPG is manufactured when crude oil is refined, or is extracted from natural gas supplies in the ground.
Paraffin	This is a technical name for the alkanes, but refers specifically to the <i>linear</i> alkanes. <i>Isoparaffin</i> refers to non-linear (branched) alkanes.
Jet fuel	A type of aviation fuel designed for use in jet engined aircraft. It is an oil-based fuel and contains additives such as antioxidants, corrosion inhibitors and icing inhibitors.

You will notice in the diagram that Sasol doesn't only produce liquid fuels, but also a variety of other chemical products. Sometimes it is the synthetic fuels themselves that are used as feedstocks to produce these chemical products. This is done through processes such as **hydrocracking** and **steamcracking**. Cracking is when heavy hydrocarbons are converted to simpler light hydrocarbons (e.g. LPG and petrol) through the breaking of C-C bonds. A heavy hydrocarbon is one that has a high number of hydrogen and carbon atoms (more solid), and a light hydrocarbon has fewer hydrogen and carbon atoms and is either a liquid or a gas.



Definition: Hydrocracking

Hydrocracking is a cracking process that is assisted by the presence of an elevated partial pressure of hydrogen gas. It produces chemical products such as ethane, LPG, isoparaffins, jet fuel and diesel.

**Definition: Steam cracking**

Steam cracking occurs under very high temperatures. During the process, a liquid or gaseous hydrocarbon is diluted with steam and then briefly heated in a furnace at a temperature of about $850^{\text{circ}}\text{C}$. Steam cracking is used to convert *ethane* to *ethylene*. Ethylene is a chemical that is needed to make plastics. Steam cracking is also used to make propylene, which is an important fuel gas.

2. Production of crude oil

Sasol obtains crude oil off the coast of Gabon (a country in West Africa) and refines this at the Natref refinery (figure 5.2). Sasol also sells liquid fuels through a number of service stations.

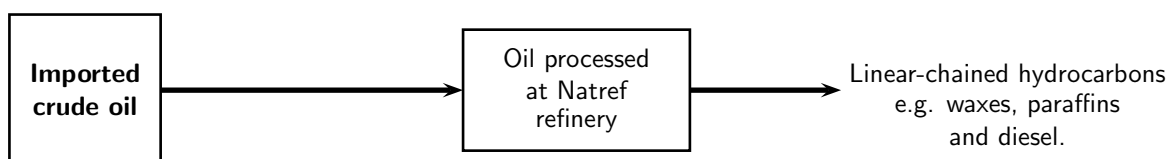


Figure 5.2: Crude oil is refined at Sasol's Natref refinery to produce liquid fuels

3. Liquid fuels from natural gas

Sasol produces natural gas in Mozambique and is expanding its 'gas to fuel' technology. The gas undergoes a complex process to produce linear-chained hydrocarbons such as waxes and paraffins (figure 5.3).

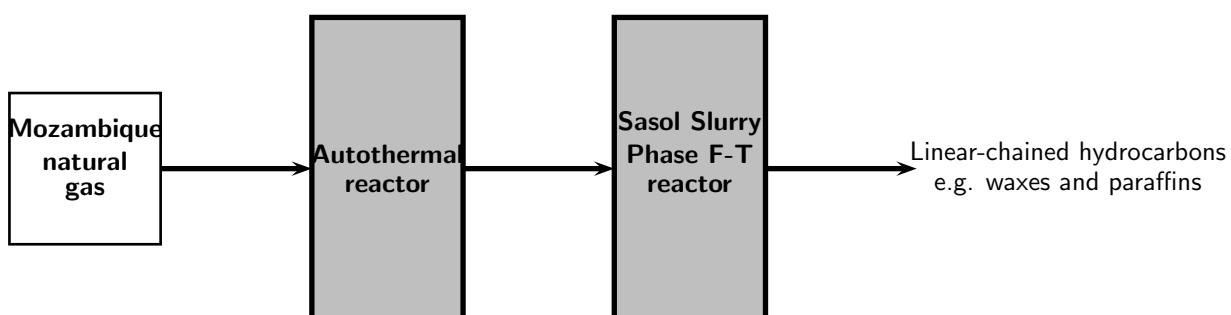


Figure 5.3: Conversion of natural gas to liquid fuels

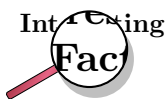
In the **autothermal reactor**, methane from natural gas reacts with steam and oxygen over an iron-based catalyst to produce a *synthesis gas*. This is a similar process to that involved in coal gasification. The oxygen is produced through the **fractional distillation of air**.

**Definition: Fractional distillation**

Fractional distillation is the separation of a mixture into its component parts, or fractions. Since air is made up of a number of gases (with the major component being nitrogen), fractional distillation can be used to separate it into these different parts.

The syngas is then passes through a **Sasol Slurry Phase Distillate (SSPD)** process. In this process, the gas is reacted at far lower temperatures than in the SAS reactors. Apart

from hard wax and candle wax, high quality diesel can also be produced in this process. Residual gas from the SSPD process is sold as pipeline gas while some of the lighter hydrocarbons are treated to produce kerosene and paraffin. Ammonia is also produced, which can be used to make fertilisers.



Sasol is a major player in the emerging Southern African natural gas industry, after investing 1.2 billion US dollars to develop onshore gas fields in central Mozambique. Sasol has been supplying natural gas from Mozambique's Temane field to customers in South Africa since 2004.



Exercise: Sasol processes

Refer to the diagrams summarising the three main Sasol processes, and use these to answer the following questions:

- Explain what is meant by each of the following terms:
 - crude oil
 - hydrocarbon
 - coal gasification
 - synthetic fuel
 - chemical feedstock
- What is diesel?
 - Describe two ways in which diesel can be produced.
- Describe one way in which lighter chemical products such as ethylene, can be produced.
- Coal and oil play an important role in Sasol's technology.
 - In the table below, summarise the similarities and differences between coal, oil and natural gas in terms of how they are formed ('origin'), their general chemical formula and whether they are solid, liquid or gas.

	Coal	Oil	Natural gas
Origin			
General chemical formula			
Solid, liquid or gas			

- In your own words, describe how coal is converted into liquid fuels.
- Explain why Sasol's 'coal to liquid fuels' technology is so important in meeting South Africa's fuel needs.
- Low grade coal is used to produce liquid fuels. What is the main use of higher grade coal in South Africa?

Activity :: Case Study : Safety issues and risk assessments

Safety issues are important to consider when dealing with industrial processes. Read the following extract that appeared in the Business report on 6th February 2006, and then discuss the questions that follow.

Cape Town - Sasol, the petrochemicals group, was likely to face prosecution on 10 charges of culpable homicide after an explosion at its Secunda plant in 2004 in which 10 people died, a Cape Town labour law specialist said on Friday. The specialist, who did not want to be named, was speaking after the inquiry into the explosion was concluded last Tuesday. It was convened by the labour department.

The evidence led at the inquiry showed a failure on the part of the company to conduct a proper risk assessment and that: Sasol failed to identify hazards associated with a high-pressure gas pipeline running through the plant, which had been shut for extensive maintenance work, in the presence of hundreds of people and numerous machines, including cranes, fitters, contractors, and welding and cutting machines. Because there had never been a risk assessment, the hazard of the high-pressure pipeline had never been identified.

Because Sasol had failed to identify the risk, it did not take any measures to warn people about it, mark the line or take precautions. There had also been inadequacy in planning the shutdown work. In the face of a barrage of criticism for the series of explosions that year, Sasol embarked on a comprehensive programme to improve safety at its operations and appointed Du Pont Safety Resources, the US safety consultancy, to benchmark the petrochemical giant's occupational health and safety performance against international best practice.

1. Explain what is meant by a 'risk assessment'.
2. Imagine that you have been asked to conduct a risk assessment of the Sasol/Lurgi gasification process. What information would you need to know in order to do this assessment?
3. In groups, discuss the importance of each of the following in ensuring the safety of workers in the chemical industry:
 - employing experienced Safety, Health and Environment personnel
 - regular training to identify hazards
 - equipment maintenance and routine checks
4. What other precautions would you add to this list to make sure that working conditions are safe?

5.2.2 Sasol and the environment

From its humble beginnings in 1950, Sasol has grown to become a major contributor towards the South African economy. Today, the industry produces more than 150 000 barrels of fuels and petrochemicals per day, and meets more than 40% of South Africa's liquid fuel requirements. In total, more than 200 fuel and chemical products are manufactured at Sasolburg and Secunda, and these products are exported to over 70 countries worldwide. This huge success is largely due to Sasol's ability to diversify its product base. The industry has also helped to provide about 170 000 jobs in South Africa, and contributes around R40 billion to the country's Gross Domestic Product (GDP).

However, despite these obvious benefits, there are always environmental costs associated with industry. Apart from the vast quantities of resources that are needed in order for the industry to operate, the production process itself produces waste products and pollutants.



Exercise: Consumption of resources

Any industry will always use up huge amounts of resources in order to function effectively, and the chemical industry is no exception. In order for an industry to operate, some of the major resources that are needed are **energy** to drive many of the processes, **water**, either as a coolant or as part of a process and **land** for mining or operations.

Refer to the data table below which shows Sasol's water use between 2002 and 2005 (*Sasol Sustainable Development Report 2005*), and answer the questions that follow.

Water use ($1000m^3$)	2002	2003	2004	2005
River water	113 722	124 179	131 309	124 301
Potable water	15 126	10 552	10 176	10 753
Total	157 617	178 439	173 319	163 203

1. Explain what is meant by 'potable' water.
2. Describe the trend in Sasol's water use that you see in the above statistics.
3. Suggest possible reasons for this trend.
4. List some of the environmental impacts of using large amounts of river water for industry.
5. Suggest ways in which these impacts could be reduced



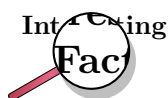
Exercise: Industry and the environment

Large amounts of gases and pollutants are released during production, and when the fuels themselves are used. Refer to the table below, which shows greenhouse gas and atmospheric pollution data for Sasol between 2002 and 2005, and then answer the questions that follow. (*Source: Sasol Sustainable Development Report 2005*)

Greenhouse gases and air pollutants (kilotonnes)	2002	2003	2004	2005
Carbon dioxide (CO_2)	57 476	62 873	66 838	60 925
Hydrogen sulfide (H_2S)	118	105	102	89
Nitrogen oxides (NO_x)	168	173	178	166
Sulfur dioxide (SO_2)	283	239	261	222

1. Draw line graphs to show how the quantity of each pollutant produced has changed between 2002 and 2005.
2. Describe what you see in the graphs, and suggest a reason for this trend.
3. Explain what is meant by each of the following terms:
 - (a) greenhouse gas
 - (b) global warming
4. Describe some of the possible effects of global warming.
5. When sulfur dioxide is present in the atmosphere, it may react with water vapour to produce *sulfuric acid*. In the same way, nitrogen dioxide and water vapour react to form *nitric acid*. These reactions in the atmosphere may cause **acid rain**. Outline some of the possible consequences of acid rain.
6. Many industries are major contributors towards environmental problems such as global warming, environmental pollution, over-use of resources and acid rain. Industries are in a difficult position: On one hand they must meet the ever increasing demands of society, and on the other, they must achieve this with as little environmental impact as possible. This is a huge challenge.

- Work in groups of 3-4 to discuss ways in which industries could be encouraged (or in some cases forced) to reduce their environmental impact.
 - Elect a spokesperson for each group, who will present your ideas to the class.
 - Are the ideas suggested by each group practical?
 - How easy or difficult do you think it would be to implement these ideas in South Africa?
-
-



Sasol is very aware of its responsibility towards creating cleaner fuels. From 1st January 2006, the South African government enforced a law to prevent lead from being added to petrol. Sasol has complied with this. One branch of Sasol, **Sasol Technology** also has a bio-diesel research and development programme focused on developing more environmentally friendly forms of diesel. One way to do this is to use renewable resources such as soybeans to make diesel. Sasol is busy investigating this new technology.

5.3 The Chloralkali Industry

The chlorine-alkali (chloralkali) industry is an important part of the chemical industry, and produces **chlorine** and **sodium hydroxide** through the electrolysis of table salt (NaCl). The main raw material is **brine** which is a saturated solution of sodium chloride (NaCl) that is obtained from natural salt deposits.

The products of this industry have a number of important uses. **Chlorine** is used to purify water, and is used as a disinfectant. It is also used in the manufacture of many every-day items such as hypochlorous acid, which is used to kill bacteria in drinking water. Chlorine is also used in paper production, antiseptics, food, insecticides, paints, petroleum products, plastics (such as polyvinyl chloride or PVC), medicines, textiles, solvents, and many other consumer products. Many chemical products such as chloroform and carbon tetrachloride also contain chlorine.

Sodium hydroxide (also known as 'caustic soda') has a number of uses, which include making soap and other cleaning agents, purifying bauxite (the ore of aluminium), making paper and making rayon (artificial silk).

5.3.1 The Industrial Production of Chlorine and Sodium Hydroxide

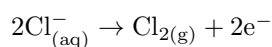
Chlorine and sodium hydroxide can be produced through a number of different reactions. However, one of the problems is that when chlorine and sodium hydroxide are produced together, the chlorine combines with the sodium hydroxide to form chlorate (ClO^-) and chloride (Cl^-) ions. This produces sodium chlorate, NaClO, a component of household bleach. To overcome this problem the chlorine and sodium hydroxide must be separated from each other so that they don't react. There are three industrial processes that have been designed to overcome this problem, and to produce chlorine and sodium hydroxide. All three methods involve **electrolytic cells** (chapter 4).

Important: Electrolytic cells

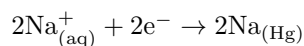
Electrolytic cells are used to transform reactants into products by using electric current. They are made up of an **electrolyte** and two electrodes, the **cathode** and the **anode**. An electrolytic cell is activated by applying an external electrical current. This creates an electrical potential across the cathode and anode, and forces a chemical reaction to take place in the electrolyte. Cations flow towards the cathode and are reduced. Anions flow to the anode and are oxidised. Two new products are formed, one product at the cathode and one at the anode.

1. The Mercury Cell

In the mercury-cell (figure 5.4), brine passes through a chamber which has a carbon electrode (the anode) suspended from the top. Mercury flows along the floor of this chamber and acts as the cathode. When an electric current is applied to the circuit, chloride ions in the electrolyte are oxidised to form chlorine gas.



At the cathode, sodium ions are reduced to sodium.



The sodium dissolves in the mercury, forming an amalgam of sodium and mercury. The amalgam is then poured into a separate vessel, where it decomposes into sodium and mercury. The sodium reacts with water in the vessel and produces sodium hydroxide (caustic soda) and hydrogen gas, while the mercury returns to the electrolytic cell to be used again.

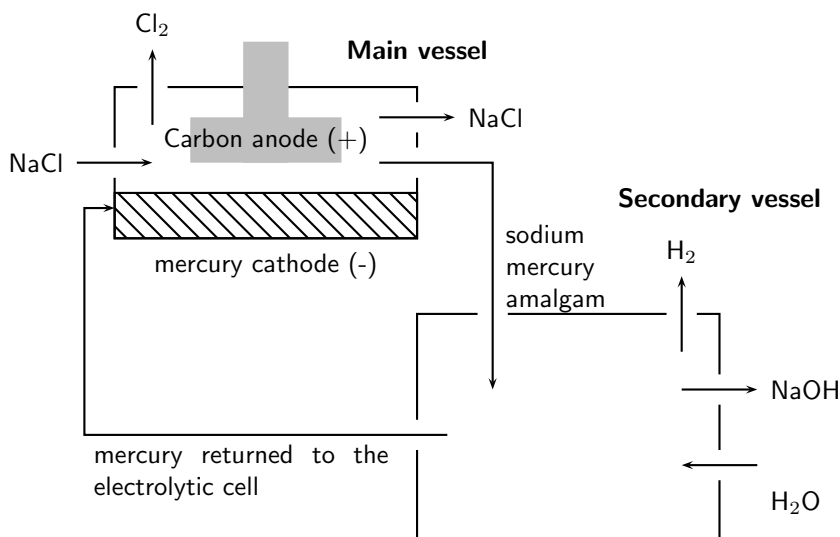
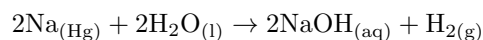


Figure 5.4: The Mercury Cell

This method, however, only produces a fraction of the chlorine and sodium hydroxide that is used by industry as it has certain disadvantages: mercury is expensive and toxic, and although it is returned to the electrolytic cell, some always escapes with the brine that has been used. The mercury reacts with the brine to form mercury(II) chloride. In the past this effluent was released into lakes and rivers, causing mercury to accumulate in fish and other animals feeding on the fish. Today, the brine is treated before it is discharged so that the environmental impact is lower.

2. The Diaphragm Cell

In the diaphragm-cell (figure 5.5), a porous diaphragm divides the electrolytic cell, which contains brine, into an anode compartment and a cathode compartment. The brine is introduced into the anode compartment and flows through the diaphragm into the cathode compartment. When an electric current passes through the brine, the salt's chlorine ions and sodium ions move to the electrodes. Chlorine gas is produced at the anode. At the cathode, sodium ions react with water, forming caustic soda and hydrogen gas. Some salt remains in the solution with the caustic soda and can be removed at a later stage.

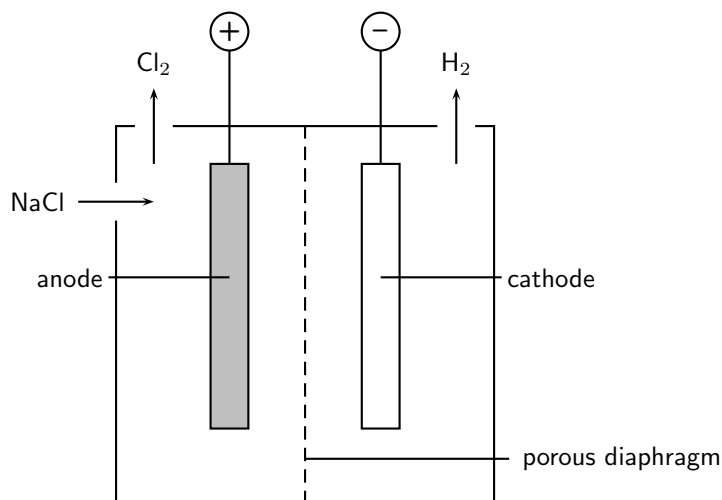
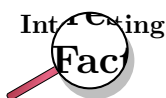


Figure 5.5: Diaphragm Cell

This method uses less energy than the mercury cell, but the sodium hydroxide is not as easily concentrated and precipitated into a useful substance.

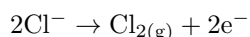


To separate the chlorine from the sodium hydroxide, the two half-cells were traditionally separated by a porous asbestos diaphragm, which needed to be replaced every two months. This was damaging to the environment, as large quantities of asbestos had to be disposed. Today, the asbestos is being replaced by other polymers which do not need to be replaced as often.

3. The Membrane Cell

The membrane cell (figure 5.6) is very similar to the diaphragm cell, and the same reactions occur. The main difference is that the two electrodes are separated by an ion-selective membrane, rather than by a diaphragm. The structure of the membrane is such that it allows cations to pass through it between compartments of the cell. It does not allow anions to pass through. This has nothing to do with the size of the pores, but rather with the charge on the ions. Brine is pumped into the anode compartment, and only the positively charged sodium ions pass into the cathode compartment, which contains pure water.

At the positively charged anode, Cl^- ions from the brine are oxidised to Cl_2 gas.



At the negatively charged cathode, hydrogen ions in the water are reduced to hydrogen gas.

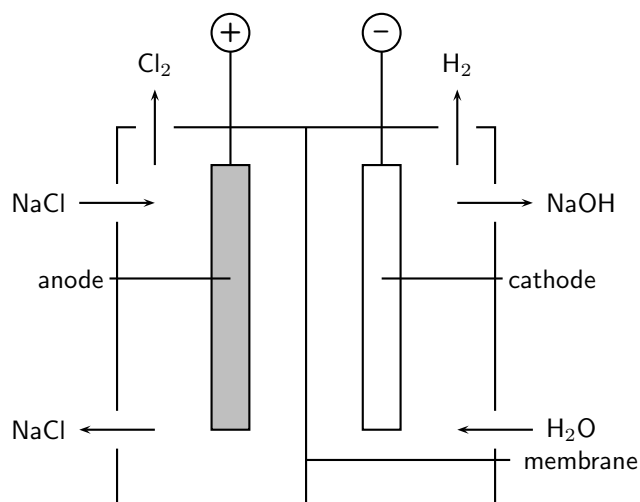
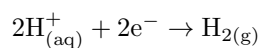
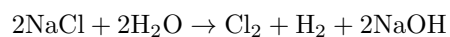


Figure 5.6: Membrane Cell



The Na^{+} ions flow through the membrane to the cathode compartment and react with the remaining hydroxide (OH^{-}) ions from the water to form sodium hydroxide (NaOH). The chloride ions cannot pass through, so the chlorine does not come into contact with the sodium hydroxide in the cathode compartment. The sodium hydroxide is removed from the cell. The overall equation is as follows:

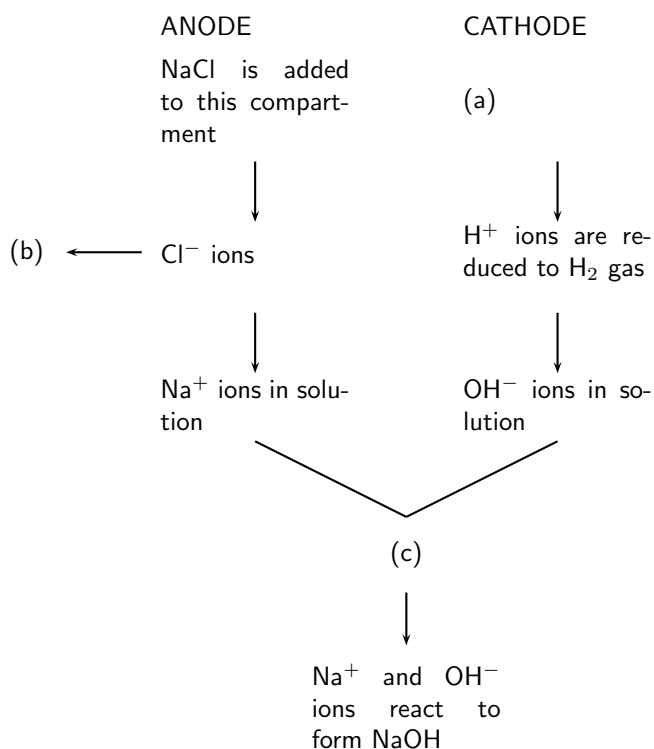


The advantage of using this method is that the sodium hydroxide that is produced is very pure because it is kept separate from the sodium chloride solution. The caustic soda therefore has very little salt contamination. The process also uses less electricity and is cheaper to operate.



Exercise: The Chloralkali industry

1. Refer to the flow diagram below which shows the reactions that take place in the membrane cell, and then answer the questions that follow.



- What liquid is present in the cathode compartment at (a)?
 - Identify the gas that is produced at (b).
 - Explain one feature of this cell that allows the Na⁺ and OH⁻ ions to react at (c).
 - Give a balanced equation for the reaction that takes place at (c).
2. Summarise what you have learnt about the three types of cells in the chloralkali industry by completing the table below:

	Mercury cell	Diaphragm cell	Membrane cell
Main raw material			
Mechanism of separating Cl ₂ and NaOH			
Anode reaction			
Cathode reaction			
Purity of NaOH produced			
Energy consumption			
Environmental impact			

5.3.2 Soaps and Detergents

Another important part of the chloralkali industry is the production of **soaps** and **detergents**. You will remember from an earlier chapter, that water has the property of *surface tension*. This means that it tends to bead up on surfaces and this slows down the wetting process and makes cleaning difficult. You can observe this property of surface tension when a drop of water falls onto a table surface. The drop holds its shape and does not spread. When cleaning, this surface tension must be reduced so that the water can spread. Chemicals that are able to do this

are called **surfactants**. Surfactants also loosen, disperse and hold particles in suspension, all of which are an important part of the cleaning process. Soap is an example of one of these surfactants. Detergents contain one or more surfactants. We will go on to look at these in more detail.



Definition: Surfactant

A surfactant is a wetting agent that lowers the surface tension of a liquid, allowing it to spread more easily.

1. Soaps

In chapter 2, a number of important biological macromolecules were discussed, including carbohydrates, proteins and nucleic acids. **Fats** are also biological macromolecules. A fat is made up of an alcohol called glycerol, attached to three fatty acids (figure 5.7). Each **fatty acid** is made up of a carboxylic acid attached to a long hydrocarbon chain. An **oil** has the same structure as a fat, but is a liquid rather than a solid. Oils are found in plants (e.g. olive oil, sunflower oil) and fats are found in animals.

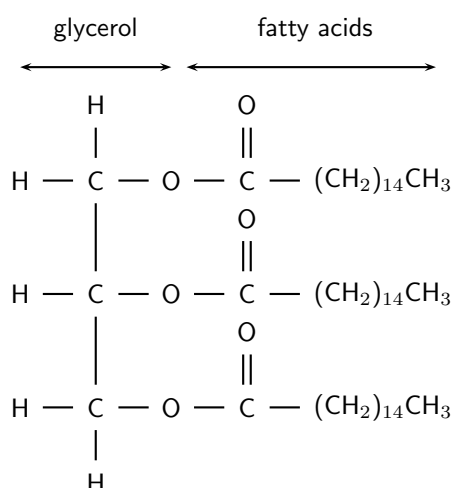
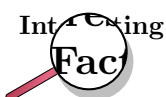


Figure 5.7: The structure of a fat, composed of an alcohol and three fatty acids

To make soap, sodium hydroxide (NaOH) or potassium hydroxide (KOH) must be added to a fat or an oil. During this reaction, the glycerol is separated from the fatty acid in the fat, and is replaced by either potassium or sodium ions (figure 5.8). Soaps are the water-soluble sodium or potassium salts of fatty acids.



Soaps can be made from either fats or oils. Beef fat is a common source of fat, and vegetable oils such as palm oil are also commonly used.

Fatty acids consist of two parts: a carboxylic acid group and a hydrocarbon chain. The hydrocarbon chain is *hydrophobic*, meaning that it is repelled by water. However, it is attracted to grease, oils and other dirt. The carboxylic acid is *hydrophilic*, meaning that it is attracted to water. Let's imagine that we have added soap to water in order to clean a dirty rugby jersey. The hydrocarbon chain will attach itself to the soil particles in the

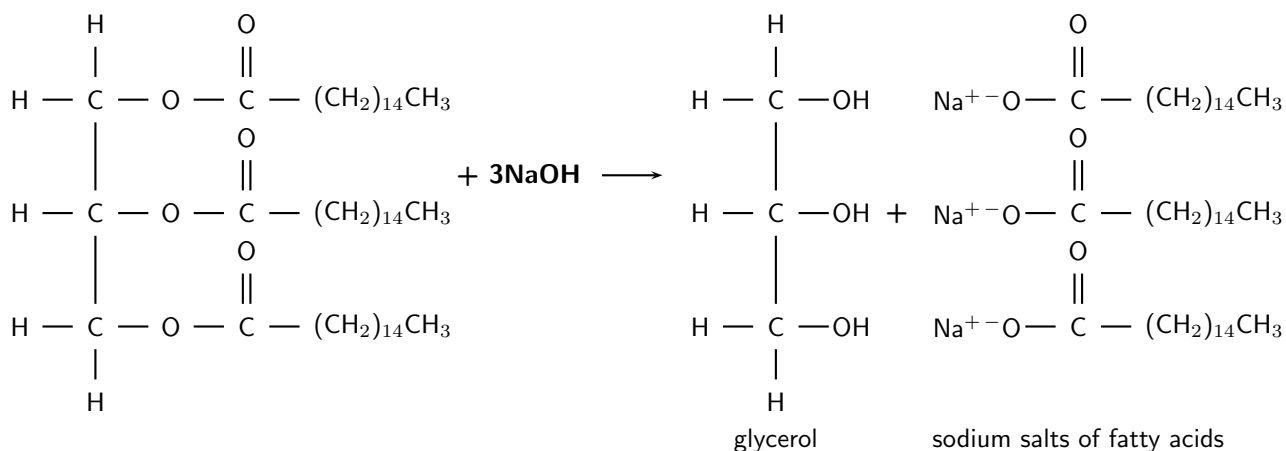


Figure 5.8: Sodium hydroxide reacts with a fat to produce glycerol and sodium salts of the fatty acids

jersey, while the carboxylic acid will be attracted to the water. In this way, the soil is pulled free of the jersey and is suspended in the water. In a washing machine or with vigorous handwashing, this suspension can be rinsed off with clean water.



Definition: Soap

Soap is a surfactant that is used with water for washing and cleaning. Soap is made by reacting a fat with either sodium hydroxide (NaOH) or potassium hydroxide (KOH).

2. Detergents



Definition: Detergent

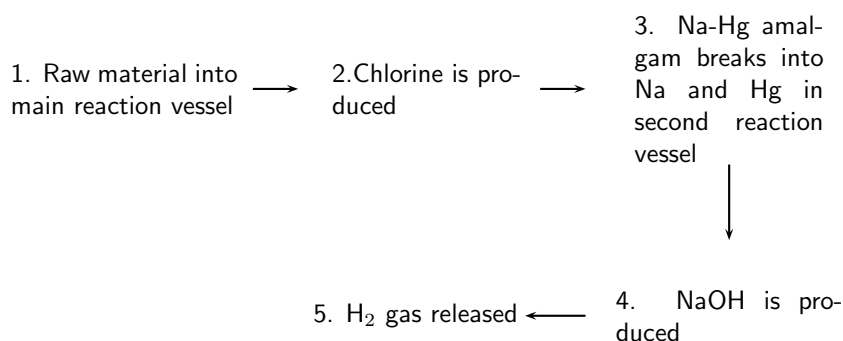
Detergents are compounds or mixtures of compounds that are used to assist cleaning. The term is often used to distinguish between soap and other chemical surfactants for cleaning.

Detergents are also cleaning products, but are composed of one or more surfactants. Depending on the type of cleaning that is needed, detergents may contain one or more of the following:

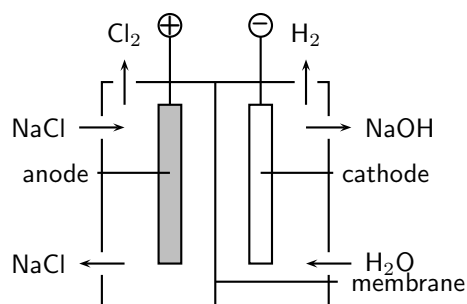
- *Abrasives* to scour a surface.
- *Oxidants* for bleaching and disinfection.
- *Enzymes* to digest proteins, fats or carbohydrates in stains. These are called *biological detergents*.



Exercise: The chloralkali industry



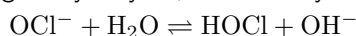
- The diagram above shows the sequence of steps that take place in the mercury cell.
 - Name the 'raw material' in step 1.
 - Give the chemical equation for the reaction that produces chlorine in step 2.
 - What other product is formed in step 2.
 - Name the reactants in step 4.
- Approximately 30 million tonnes of chlorine are used throughout the world annually. Chlorine is produced industrially by the electrolysis of brine. The diagram represents a membrane cell used in the production of Cl₂ gas.



- What ions are present in the electrolyte in the left hand compartment of the cell?
- Give the equation for the reaction that takes place at the anode.
- Give the equation for the reaction that takes place at the cathode.
- What ion passes through the membrane while these reactions are taking place?

Chlorine is used to purify drinking water and swimming pool water. The substance responsible for this process is the weak acid, hypochlorous acid (HOCl).

- One way of putting HOCl into a pool is to bubble chlorine gas through the water. Give an equation showing how bubbling Cl₂(g) through water produces HOCl.
- A common way of treating pool water is by adding 'granular chlorine'. Granular chlorine consists of the salt calcium hypochlorite, Ca(OCl)₂. Give an equation showing how this salt dissolves in water. Indicate the phase of each substance in the equation.
- The OCl⁻ ion undergoes hydrolysis, as shown by the following equation:



Will the addition of granular chlorine to pure water make the water acidic, basic or will it remain neutral? Briefly explain your answer.

(IEB Paper 2, 2003)

5.4 The Fertiliser Industry

5.4.1 The value of nutrients

Nutrients are very important for life to exist. An **essential nutrient** is any chemical element that is needed for a plant to be able to grow from a seed and complete its life cycle. The same is true for animals. A *macronutrient* is one that is required in large quantities by the plant or animal, while a *micronutrient* is one that only needs to be present in small amounts for a plant or an animal to function properly.



Definition: Nutrient

A nutrient is a substance that is used in an organism's metabolism or physiology and which must be taken in from the environment.

In plants, the macronutrients include carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P) and potassium (K). The source of each of these nutrients for plants, and their function, is summarised in table 5.1. Examples of micronutrients in plants include iron, chlorine, copper and zinc.

Table 5.1: The source and function of the macronutrients in plants

Nutrient	Source	Function
Carbon	Carbon dioxide in the air	Component of organic molecules such as carbohydrates, lipids and proteins
Hydrogen	Water from the soil	Component of organic molecules
Oxygen	Water from the soil	Component of organic molecules
Nitrogen	Nitrogen compounds in the soil	Part of plant proteins and chlorophyll. Also boosts plant growth.
Phosphorus	Phosphates in the soil	Needed for photosynthesis, blooming and root growth
Potassium	Soil	Building proteins, part of chlorophyll and reduces diseases in plants

Animals need similar nutrients in order to survive. However since animals can't photosynthesise, they rely on plants to supply them with the nutrients they need. Think for example of the human diet. We can't make our own food and so we either need to eat vegetables, fruits and seeds (all of which are direct plant products) or the meat of other animals which would have fed on plants during their life. So most of the nutrients that animals need are obtained either directly or indirectly from plants. Table 5.2 summarises the functions of some of the macronutrients in animals.

Micronutrients also play an important function in animals. Iron for example, is found in haemoglobin, the blood pigment that is responsible for transporting oxygen to all the cells in the body.

Nutrients then, are essential for the survival of life. Importantly, obtaining nutrients starts with plants, which are able either to photosynthesise or to absorb the required nutrients from the soil. It is important therefore that plants are always able to access the nutrients that they need so that they will grow and provide food for other forms of life.

5.4.2 The Role of fertilisers

Plants are only able to absorb soil nutrients in a particular form. Nitrogen for example, is absorbed as **nitrates**, while phosphorus is absorbed as **phosphates**. The **nitrogen cycle** (Grade

Table 5.2: The functions of animal macronutrients

Nutrient	Function
Carbon	Component of organic compounds
Hydrogen	Component of organic compounds
Oxygen	Component of organic compounds
Nitrogen	Component of nucleic acids and proteins
Phosphorus	Component of nucleic acids and phospholipids
Potassium	Helps in coordination and regulating the water balance in the body

10) describes the process that is involved in converting atmospheric nitrogen into a form that can be used by plants.

However, all these natural processes of maintaining soil nutrients take a long time. As populations grow and the demand for food increases, there is more and more strain on the land to produce food. Often, cultivation practices don't give the soil enough time to recover and to replace the nutrients that have been lost. Today, **fertilisers** play a very important role in restoring soil nutrients so that crop yields stay high. Some of these fertilisers are **organic** (e.g. compost, manure and fishmeal), which means that they started off as part of something living. Compost for example is often made up of things like vegetable peels and other organic remains that have been thrown away. Others are **inorganic** and can be made industrially. The advantage of these commercial fertilisers is that the nutrients are in a form that can be absorbed immediately by the plant.



Definition: Fertiliser

A fertiliser is a compound that is given to a plant to promote growth. Fertilisers usually provide the three major plant nutrients and most are applied via the soil so that the nutrients are absorbed by plants through their roots.

When you buy fertilisers from the shop, you will see three numbers on the back of the packet e.g. 18-24-6. These numbers are called the **NPK ratio**, and they give the percentage of nitrogen, phosphorus and potassium in that fertiliser. Depending on the types of plants you are growing, and the way in which you would like them to grow, you may need to use a fertiliser with a slightly different ratio. If you want to encourage root growth in your plant for example, you might choose a fertiliser with a greater amount of phosphorus. Look at the table below, which gives an idea of the amounts of nitrogen, phosphorus and potassium there are in different types of fertilisers. Fertilisers also provide other nutrients such as calcium, sulfur and magnesium.

Table 5.3: Common grades of some fertiliser materials

Description	Grade (NPK %)
Ammonium nitrate	34-0-0
Urea	46-0-0
Bone Meal	4-21-1
Seaweed	1-1-5
Starter fertilisers	18-24-6
Equal NPK fertilisers	12-12-12
High N, low P and medium K fertilisers	25-5-15

5.4.3 The Industrial Production of Fertilisers

The industrial production of fertilisers may involve several processes.

1. Nitrogen fertilisers

Making **nitrogen fertilisers** involves producing *ammonia*, which is then reacted with *oxygen* to produce *nitric acid*. Nitric acid is used to acidify phosphate rock to produce nitrogen fertilisers. The flow diagram below illustrates the processes that are involved. Each of these steps will be examined in more detail.

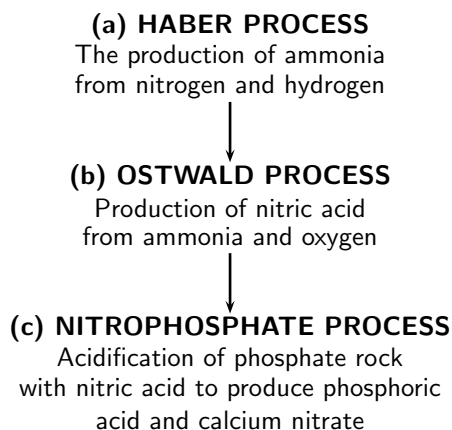
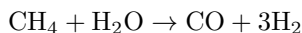


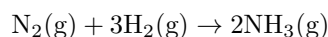
Figure 5.9: Flow diagram showing steps in the production of nitrogen fertilisers

(a) The Haber Process

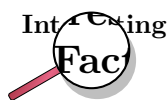
The Haber process involves the reaction of nitrogen and hydrogen to produce ammonia. **Nitrogen** is produced through the **fractional distillation** of air. Fractional distillation is the separation of a mixture (remember that air is a mixture of different gases) into its component parts through various methods. **Hydrogen** can be produced through **steam reforming**. In this process, a hydrocarbon such as methane reacts with water to form carbon monoxide and hydrogen according to the following equation:



Nitrogen and hydrogen are then used in the Haber process. The equation for the Haber process is:



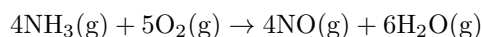
(The reaction takes place in the presence of an iron (Fe) catalyst under conditions of 200 atmospheres (atm) and 450-500 degrees Celsius)



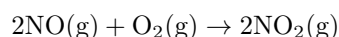
The Haber process developed in the early 20th century, before the start of World War 1. Before this, other sources of nitrogen for fertilisers had included saltpeter (NaNO_3) from Chile and guano. Guano is the droppings of seabirds, bats and seals. By the 20th century, a number of methods had been developed to 'fix' atmospheric nitrogen. One of these was the Haber process, and it advanced through the work of two German men, Fritz Haber and Karl Bosch (The process is sometimes also referred to as the 'Haber-Bosch process'). They worked out what the best conditions were in order to get a high yield of ammonia, and found these to be high temperature and high pressure. They also experimented with different catalysts to see which worked best in that reaction. During World War 1, the ammonia that was produced through the Haber process was used to make explosives. One of the advantages for Germany was that, having perfected the Haber process, they did not need to rely on other countries for the chemicals that they needed to make them.

(b) The Ostwald Process

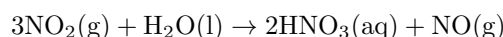
The Ostwald process is used to produce nitric acid from ammonia. Nitric acid can then be used in reactions that produce fertilisers. Ammonia is converted to nitric acid in two stages. First, it is oxidised by heating with oxygen in the presence of a platinum catalyst to form nitric oxide and water. This step is strongly exothermic, making it a useful heat source.



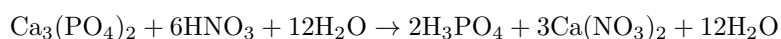
Stage two, which combines two reaction steps, is carried out in the presence of water. Initially nitric oxide is oxidised again to yield nitrogen dioxide:



This gas is then absorbed by the water to produce nitric acid. Nitric oxide is also a product of this reaction. The nitric oxide (NO) is recycled, and the acid is concentrated to the required strength.

**(c) The Nitrophosphate Process**

The nitrophosphate process involves acidifying phosphate rock with nitric acid to produce a mixture of phosphoric acid and calcium nitrate:



When calcium nitrate and phosphoric acid react with ammonia, a compound fertiliser is produced.



If potassium chloride or potassium sulphate is added, the result will be NPK fertiliser.

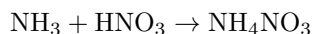
(d) Other nitrogen fertilisers

- Urea ((NH₂)₂CO) is a nitrogen-containing chemical product which is produced on a large scale worldwide. Urea has the highest nitrogen content of all solid nitrogenous fertilisers in common use (46.4%) and is produced by reacting ammonia with carbon dioxide.

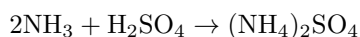
Two reactions are involved in producing urea:

- $2\text{NH}_3 + \text{CO}_2 \rightarrow \text{H}_2\text{N} - \text{COONH}_4$
- $\text{H}_2\text{N} - \text{COONH}_4 \rightarrow (\text{NH}_2)_2\text{CO} + \text{H}_2\text{O}$

- Other common fertilisers are ammonium nitrate and ammonium sulphate. Ammonium nitrate is formed by reacting ammonia with nitric acid.



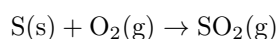
Ammonium sulphate is formed by reacting ammonia with sulphuric acid.

**2. Phosphate fertilisers**

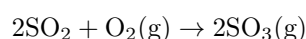
The production of phosphate fertilisers also involves a number of processes. The first is the production of sulfuric acid through the **contact process**. Sulfuric acid is then used in a reaction that produces phosphoric acid. Phosphoric acid can then be reacted with phosphate rock to produce triple superphosphates.

(a) The production of sulfuric acid

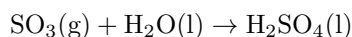
Sulfuric acid is produced from sulfur, oxygen and water through the contact process. In the first step, sulfur is burned to produce sulfur dioxide.



This is then oxidised to sulfur trioxide using oxygen in the presence of a vanadium(V) oxide catalyst.

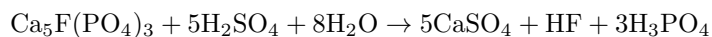


Finally the sulfur trioxide is treated with water to produce 98-99% sulfuric acid.



(b) *The production of phosphoric acid*

The next step in the production of phosphate fertiliser is the reaction of sulfuric acid with phosphate rock to produce phosphoric acid (H_3PO_4). In this example, the phosphate rock is fluorapatite ($\text{Ca}_5\text{F}(\text{PO}_4)_3$).



(c) *The production of phosphates and superphosphates*

When concentrated phosphoric acid reacts with ground phosphate rock, triple superphosphate is produced.



3. Potassium

Potassium is obtained from **potash**, an impure form of potassium carbonate (K_2CO_3). Other potassium salts (e.g. KCl AND K_2O) are also sometimes included in fertilisers.

5.4.4 Fertilisers and the Environment: Eutrophication

Eutrophication is the enrichment of an ecosystem with chemical nutrients, normally by compounds that contain nitrogen or phosphorus. Eutrophication is considered a form of pollution because it promotes plant growth, favoring certain species over others. In aquatic environments, the rapid growth of certain types of plants can disrupt the normal functioning of an ecosystem, causing a variety of problems. Human society is impacted as well because eutrophication can decrease the resource value of rivers, lakes, and estuaries making recreational activities less enjoyable. Health-related problems can also occur if eutrophic conditions interfere with the treatment of drinking water.



Definition: Eutrophication

Eutrophication refers to an increase in chemical nutrients in an ecosystem. These chemical nutrients usually contain nitrogen or phosphorus.

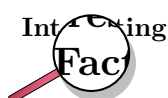
In some cases, eutrophication can be a natural process that occurs very slowly over time. However, it can also be accelerated by certain human activities. Agricultural runoff, when excess fertilisers are washed off fields and into water, and sewage are two of the major causes of eutrophication. There are a number of impacts of eutrophication.

- *A decrease in biodiversity (the number of plant and animal species in an ecosystem)*

When a system is enriched with nitrogen, plant growth is rapid. When the number of plants increases in an aquatic system, they can block light from reaching deeper. Plants also consume oxygen for respiration, and if the oxygen content of the water decreases too much, this can cause other organisms such as fish to die.

- *Toxicity*

Sometimes, the plants that flourish during eutrophication can be toxic and may accumulate in the food chain.



South Africa's Department of Water Affairs and Forestry has a 'National Eutrophication Monitoring Programme' which was set up to monitor eutrophication in impoundments such as dams, where no monitoring was taking place.

Despite the impacts, there are a number of ways of preventing eutrophication from taking place. **Cleanup measures** can directly remove the excess nutrients such as nitrogen and phosphorus from the water. Creating **buffer zones** near farms, roads and rivers can also help. These act as filters and cause nutrients and sediments to be deposited there instead of in the aquatic system. **Laws** relating to the treatment and discharge of sewage can also help to control eutrophication. A final possible intervention is **nitrogen testing and modeling**. By assessing exactly how much fertiliser is needed by crops and other plants, farmers can make sure that they only apply just enough fertiliser. This means that there is no excess to run off into neighbouring streams during rain. There is also a cost benefit for the farmer.

Activity :: Discussion : Dealing with the consequences of eutrophication

In many cases, the damage from eutrophication is already done. In groups, do the following:

1. List all the possible consequences of eutrophication that you can think of.
 2. Suggest ways to solve these problems, that arise because of eutrophication.
-
-



Exercise: Chemical industry: Fertilisers

Why we need fertilisers

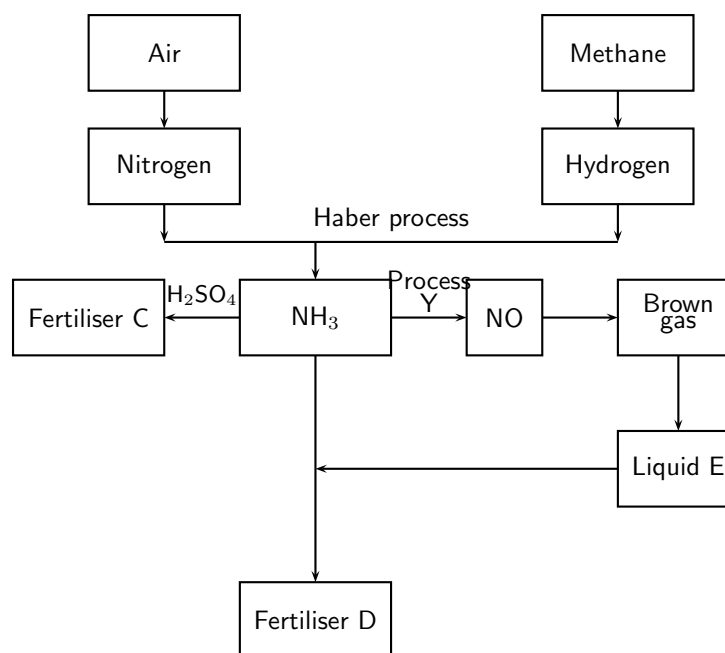
There is likely to be a gap between food production and demand in several parts of the world by 2020. Demand is influenced by population growth and urbanisation, as well as income levels and changes in dietary preferences.

The facts are as follows:

- There is an increasing world population to feed
- Most soils in the world used for large-scale, intensive production of crops lack the necessary nutrients for the crops

Conclusion: Fertilisers are needed!

The flow diagram below shows the main steps in the industrial preparation of two important solid fertilisers.



1. Write down the balanced chemical equation for the formation of the brown gas.
2. Write down the name of process Y.
3. Write down the chemical formula of liquid E.
4. Write down the chemical formulae of fertilisers C and D respectively.

The following extract comes from an article on fertilisers:

*A world without food for its people
A world with an environment poisoned through the actions of man
Are two contributing factors towards a disaster scenario.*

5. Write down THREE ways in which the use of fertilisers poisons the environment.

5.5 Electrochemistry and batteries

You will remember from chapter 4 that a **galvanic** cell (also known as a *voltaic* cell) is a type of electrochemical cell where a chemical reaction produces electrical energy. The **electromotive force** (emf) of a galvanic cell is the difference in voltage between the two half cells that make it up. Galvanic cells have a number of applications, but one of the most important is their use in **batteries**. You will know from your own experience that we use batteries in a number of ways, including cars, torches, sound systems and cellphones to name just a few.

5.5.1 How batteries work

A battery is a device in which **chemical energy** is directly converted to **electrical energy**. It consists of one or more voltaic cells, each of which is made up of two half cells that are connected in series by a conductive electrolyte. The voltaic cells are connected in series in a battery. Each cell has a positive electrode (cathode), and a negative electrode (anode). These do not touch each other but are immersed in a solid or liquid electrolyte.

Each half cell has a net electromotive force (emf) or voltage. The voltage of the battery is the difference between the voltages of the half-cells. This potential difference between the two half cells is what causes an electric current to flow.

Batteries are usually divided into two broad classes:

- *Primary batteries* irreversibly transform chemical energy to electrical energy. Once the supply of reactants has been used up, the battery can't be used any more.
- *Secondary batteries* can be recharged, in other words, their chemical reactions can be reversed if electrical energy is supplied to the cell. Through this process, the cell returns to its original state. Secondary batteries can't be recharged forever because there is a gradual loss of the active materials and electrolyte. Internal corrosion can also take place.

5.5.2 Battery capacity and energy

The **capacity** of a battery, in other words its ability to produce an electric charge, depends on a number of factors. These include:

- **Chemical reactions**

The chemical reactions that take place in each of a battery's half cells will affect the voltage across the cell, and therefore also its capacity. For example, nickel-cadmium (NiCd) cells measure about 1.2 V, and alkaline and carbon-zinc cells both measure about 1.5 V. However, in other cells such as Lithium cells, the changes in electrochemical potential are much higher because of the reactions of lithium compounds, and so lithium cells can produce as much as 3 volts or more. The concentration of the chemicals that are involved will also affect a battery's capacity. The higher the concentration of the chemicals, the greater the capacity of the battery.

- **Quantity of electrolyte and electrode material in cell**

The greater the amount of electrolyte in the cell, the greater its capacity. In other words, even if the chemistry in two cells is the same, a larger cell will have a greater capacity than a small one. Also, the greater the surface area of the electrodes, the greater will be the capacity of the cell.

- **Discharge conditions**

A unit called an **Ampere hour** (Ah) is used to describe how long a battery will last. An ampere hour (more commonly known as an **amp hour**) is the amount of electric charge that is transferred by a current of one ampere for one hour. Battery manufacturers use a standard method to rate their batteries. So, for example, a 100 Ah battery will provide a current of 5 A for a period of 20 hours at room temperature. The capacity of the battery will depend on the rate at which it is discharged or used. If a 100 Ah battery is discharged at 50 A (instead of 5 A), the capacity will be *lower* than expected and the battery will run out *before* the expected 2 hours.

The relationship between the current, discharge time and capacity of a battery is expressed by **Peukert's law**:

$$C_p = I^k t$$

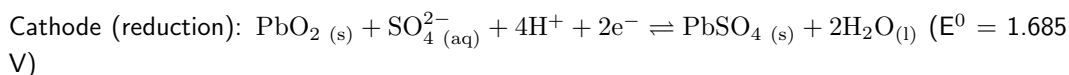
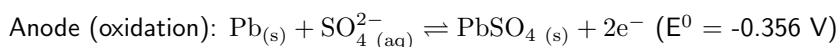
In the equation, ' C_p ' represents the battery's capacity (Ah), I is the discharge current (A), k is the Peukert constant and t is the time of discharge (hours).

5.5.3 Lead-acid batteries

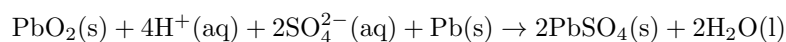
In a **lead-acid battery**, each cell consists of electrodes of lead (Pb) and lead (IV) oxide (PbO₂) in an electrolyte of sulfuric acid (H₂SO₄). When the battery discharges, both electrodes turn

into lead (II) sulphate (PbSO_4) and the electrolyte loses sulfuric acid to become mostly water.

The chemical half reactions that take place at the anode and cathode when the battery is **discharging** are as follows:



The overall reaction is as follows:



The emf of the cell is calculated as follows:

$$\text{EMF} = E(\text{cathode}) - E(\text{anode})$$

$$\text{EMF} = +1.685 \text{ V} - (-0.356 \text{ V})$$

$$\text{EMF} = +2.041 \text{ V}$$

Since most batteries consist of six cells, the total voltage of the battery is approximately 12 V.

One of the important things about a lead-acid battery is that it can be **recharged**. The recharge reactions are the *reverse* of those when the battery is discharging.

The lead-acid battery is made up of a number of *plates* that maximise the surface area on which chemical reactions can take place. Each plate is a rectangular grid, with a series of holes in it. The holes are filled with a mixture of lead and sulfuric acid. This paste is pressed into the holes and the plates are then stacked together, with suitable separators between them. They are then placed in the battery container, after which acid is added (figure 5.10).

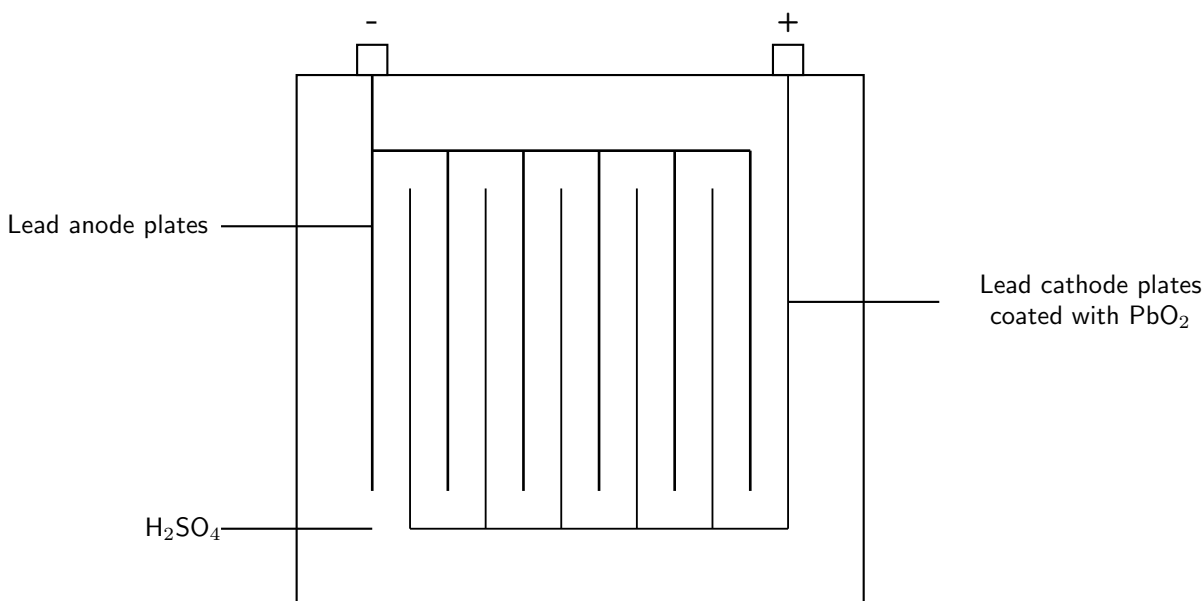


Figure 5.10: A lead-acid battery

Lead-acid batteries have a number of applications. They can supply high surge currents, are relatively cheap, have a long shelf life and can be recharged. They are ideal for use in cars,

where they provide the high current that is needed by the starter motor. They are also used in forklifts and as standby power sources in telecommunication facilities, generating stations and computer data centres. One of the disadvantages of this type of battery is that the battery's lead must be recycled so that the environment doesn't become contaminated. Also, sometimes when the battery is charging, hydrogen gas is generated at the cathode and this can cause a small explosion if the gas comes into contact with a spark.

5.5.4 The zinc-carbon dry cell

A simplified diagram of a zinc-carbon cell is shown in figure 5.11.

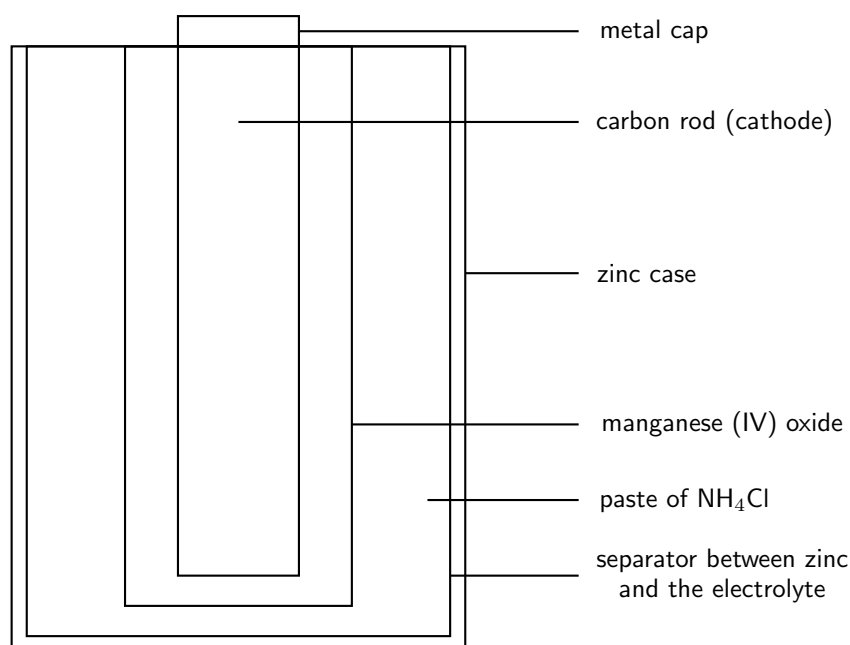
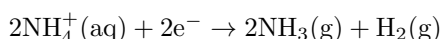


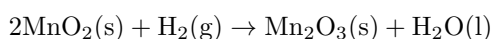
Figure 5.11: A zinc-carbon dry cell

A zinc-carbon cell is made up of an outer zinc container, which acts as the **anode**. The **cathode** is the central carbon rod, surrounded by a mixture of carbon and manganese (IV) oxide (MnO_2). The **electrolyte** is a paste of ammonium chloride (NH_4Cl). A fibrous fabric separates the two electrodes, and a brass pin in the centre of the cell conducts electricity to the outside circuit.

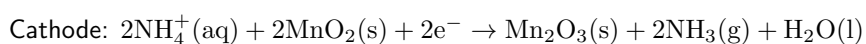
The paste of ammonium chloride reacts according to the following half-reaction:



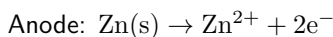
The manganese(IV) oxide in the cell removes the hydrogen produced above, according to the following reaction:



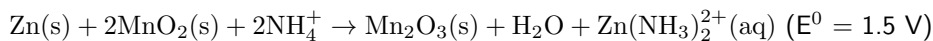
The combined result of these two reactions can be represented by the following half reaction, which takes place at the cathode:



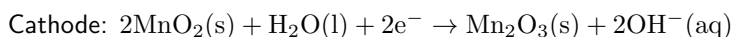
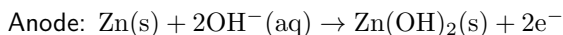
The anode half reaction is as follows:



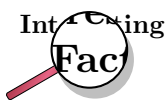
The overall equation for the cell is:



Alkaline batteries are almost the same as zinc-carbon batteries, except that the electrolyte is potassium hydroxide (KOH), rather than ammonium chloride. The two half reactions in an alkaline battery are as follows:



Zinc-carbon and alkaline batteries are cheap primary batteries and are therefore very useful in appliances such as remote controls, torches and radios where the power drain is not too high. The disadvantages are that these batteries can't be recycled and can leak. They also have a short shelf life. Alkaline batteries last longer than zinc-carbon batteries.



The idea behind today's common 'battery' was created by Georges Leclanche in France in the 1860's. The anode was a zinc and mercury alloyed rod, the cathode was a porous cup containing crushed MnO_2 . A carbon rod was inserted into this cup. The electrolyte was a liquid solution of ammonium chloride, and the cell was therefore called a *wet cell*. This was replaced by the *dry cell* in the 1880's. In the dry cell, the zinc can which contains the electrolyte, has become the anode, and the electrolyte is a paste rather than a liquid.

5.5.5 Environmental considerations

While batteries are very convenient to use, they can cause a lot of damage to the environment. They use lots of valuable resources as well as some potentially hazardous chemicals such as lead, mercury and cadmium. Attempts are now being made to recycle the different parts of batteries so that they are not disposed of in the environment, where they could get into water supplies, rivers and other ecosystems.

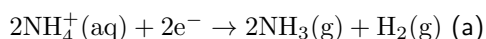


Exercise: Electrochemistry and batteries

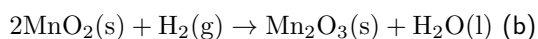
A dry cell, as shown in the diagram below, does not contain a liquid electrolyte. The electrolyte in a typical zinc-carbon cell is a moist paste of ammonium chloride and zinc chloride.

(NOTE TO SELF: Insert diagram)

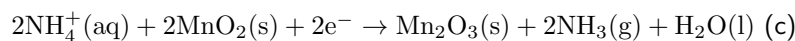
The paste of ammonium chloride reacts according to the following half-reaction:



Manganese(IV) oxide is included in the cell to remove the hydrogen produced during half-reaction (a), according to the following reaction:



The combined result of these two half-reactions can be represented by the following half reaction:



1. Explain why it is important that the hydrogen produced in half-reaction (a) is removed by the manganese(IV) oxide.
In a zinc-carbon cell, such as the one above, half-reaction (c) and the half-reaction that takes place in the Zn/Zn^{2+} half-cell, produce an emf of 1,5 V under standard conditions.
2. Write down the half-reaction occurring at the anode.
3. Write down the net ionic equation occurring in the zinc-carbon cell.
4. Calculate the reduction potential for the cathode half-reaction.
5. When in use the zinc casing of the dry cell becomes thinner, because it is oxidised. When not in use, it still corrodes. Give a reason for the latter observation.
6. Dry cells are generally discarded when 'flat'. Why is the carbon rod the most useful part of the cell, even when the cell is flat?

(DoE Exemplar Paper 2, 2007)

5.6 Summary

- The growth of South Africa's **chemical industry** was largely because of the mines, which needed explosives for their operations. One of South Africa's major chemical companies is **Sasol**. Other important chemical industries in the country are the **chloralkali** and **fertiliser** industries.
- All countries need energy resources such as oil and natural gas. Since South Africa doesn't have either of these resources, Sasol technology has developed to convert **coal** into liquid fuels.
- Sasol has three main operation focus areas: Firstly, the conversion of **coal to liquid fuel**, secondly the production and **refinement of crude oil** which has been imported, and thirdly the production of **liquid fuels from natural gas**.
- The conversion of coal to liquid fuels involves a **Sasol/Lurgi gasification** process, followed by the conversion of this synthesis gas into a range of hydrocarbons, using the **Fischer-Tropsch technology** in **SAS reactors**.
- Heavy hydrocarbons can be converted into light hydrocarbons through a process called **cracking**. Common forms of cracking are **hydrocracking** and **steam cracking**.
- With regard to crude oil, Sasol imports crude oil from Gabon and then refines this at the **Natref refinery**.
- Gas from Mozambique can be used to produce liquid fuels, through two processes: First, the gas must pass through an autothermal reactor to produce a synthesis gas. Secondly, this synthesis gas is passed through a **Sasol Slurry Phase Distillate** process to convert the gas to hydrocarbons.
- All industries have an impact on the **environment** through the consumption of natural resources such as water, and through the production of pollution gases such as carbon dioxide, hydrogen sulfides, nitrogen oxides and others.
- The **chloralkali industry** produces **chlorine** and **sodium hydroxide**. The main raw material is **brine** (NaCl).

- In industry, **electrolytic cells** are used to split the sodium chloride into its component ions to produce chlorine and sodium hydroxide. One of the challenges in this process is to keep the products of the electrolytic reaction (i.e. the chlorine and the sodium hydroxide) separate so that they don't react with each other. Specially designed electrolytic cells are needed to do this.
- There are three types of electrolytic cells that are used in this process: **mercury cell**, the **diaphragm cell** and the **membrane cell**.
- The **mercury cell** consists of two reaction vessels. The first reaction vessel contains a mercury cathode and a carbon anode. An electric current passed through the brine produces Cl^- and Na^+ ions. The Cl^- ions are oxidised to form chlorine gas at the anode. Na^+ ions combine with the mercury cathode to form a sodium-mercury amalgam. The sodium-mercury amalgam passes into the second reaction vessel containing water, where the Na^+ ions react with hydroxide ions from the water. Sodium hydroxide is the product of this reaction.
- One of the **environmental impacts** of using this type of cell, is the use of **mercury**, which is highly toxic.
- In the **diaphragm cell**, a porous diaphragm separates the anode and the cathode compartments. Chloride ions are oxidised to chlorine gas at the anode, while sodium ions produced at the cathode react with water to produce sodium hydroxide.
- The **membrane cell** is very similar to the diaphragm cell, except that the anode and cathode compartments are separated by an **ion-selective membrane** rather than by a diaphragm. Brine is only pumped into the anode compartment. Positive sodium ions pass through the membrane into the cathode compartment, which contains water. As with the other two cells, chlorine gas is produced at the anode and sodium hydroxide at the cathode.
- One use of sodium hydroxide is in the production of **soaps and detergents**, and so this is another important part of the chloralkali industry.
- To make soap, sodium hydroxide or potassium hydroxide react with a fat or an oil. In the reaction, the sodium or potassium ions replace the alcohol in the fat or oil. The product, a **sodium or potassium salt of a fatty acid**, is what soap is made of.
- The fatty acids in soap have a **hydrophilic** and a **hydrophobic** part in each molecule, and this helps to loosen dirt and clean items.
- **Detergents** are also cleaning products, but are made up of a mixture of compounds. They may also have other components added to them to give certain characteristics. Some of these additives may be abrasives, oxidants or enzymes.
- The **fertiliser industry** is another important chemical industry.
- All plants need certain **macronutrients** (e.g. carbon, hydrogen, oxygen, potassium, nitrogen and phosphorus) and **micronutrients** (e.g. iron, chlorine, copper and zinc) in order to survive. Fertilisers provide these nutrients.
- In plants, most nutrients are obtained from the atmosphere or from the soil.
- Animals also need similar nutrients, but they obtain most of these directly from plants or plant products. They may also obtain them from other animals, which may have fed on plants during their life.
- The fertiliser industry is very important in ensuring that plants and crops receive the correct nutrients in the correct quantities to ensure maximum growth.
- **Nitrogen fertilisers** can be produced industrially using a number of chemical processes: The **Haber process** reacts nitrogen and hydrogen to produce **ammonia**; the **Ostwald process** reacts oxygen and ammonia to produce **nitric acid**; the **nitrophosphate process** reacts nitric acid with phosphate rock to produce compound fertilisers.

- **Phosphate fertilisers** are also produced through a series of reactions. The **contact process** produces **sulfuric acid**. Sulfuric acid then reacts with phosphate rock to produce **phosphoric acid**, after which phosphoric acid reacts with ground phosphate rock to produce fertilisers such as **triple superphosphate**.
- Potassium is obtained from **potash**.
- Fertilisers can have a damaging effect on the environment when they are present in high quantities in ecosystems. They can lead to **eutrophication**. A number of preventative actions can be taken to reduce these impacts.
- Another important part of the chemical industry is the production of **batteries**.
- A battery is a device that changes chemical energy into electrical energy.
- A battery consists of one or more **voltaic cells**, each of which is made up of two half cells that are connected in series by a conductive electrolyte. Each half cell has a net electromotive force (emf) or voltage. The net voltage of the battery is the difference between the voltages of the half-cells. This potential difference between the two half cells is what causes an electric current to flow.
- A **primary battery** cannot be recharged, but a **secondary battery** can be recharged.
- The **capacity** of a battery depends on the **chemical reactions** in the cells, the **quantity of electrolyte and electrode material** in the cell, and the **discharge conditions** of the battery.
- The relationship between the current, discharge time and capacity of a battery is expressed by **Peukert's law**:

$$C_p = I^k t$$

In the equation, ' C_p ' represents the battery's capacity (Ah), I is the discharge current (A), k is the Peukert constant and t is the time of discharge (hours).

- Two common types of batteries are **lead-acid batteries** and the **zinc-carbon dry cell**.
- In a **lead-acid battery**, each cell consists of electrodes of lead (Pb) and lead (IV) oxide (PbO₂) in an electrolyte of sulfuric acid (H₂SO₄). When the battery discharges, both electrodes turn into lead (II) sulphate (PbSO₄) and the electrolyte loses sulfuric acid to become mostly water.
- A **zinc-carbon cell** is made up of an outer zinc container, which acts as the **anode**. The cathode is the central carbon rod, surrounded by a mixture of carbon and manganese (IV) oxide (MnO₂). The electrolyte is a paste of ammonium chloride (NH₄Cl). A fibrous fabric separates the two electrodes, and a brass pin in the centre of the cell conducts electricity to the outside circuit.
- Despite their many advantages, batteries are made of potentially toxic materials and can be damaging to the **environment**.



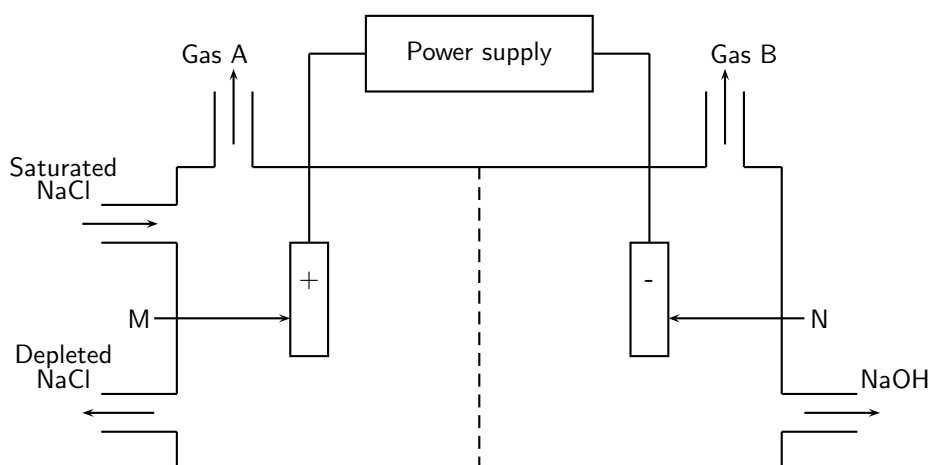
Exercise: Summary Exercise

1. Give one word or term for each of the following descriptions:
 - (a) A solid organic compound that can be used to produce liquid fuels.
 - (b) The process used to convert heavy hydrocarbons into light hydrocarbons.
 - (c) The process of separating nitrogen from liquid air.
 - (d) The main raw material in the chloralkali industry.
 - (e) A compound given to a plant to promote growth.

- (f) An electrolyte used in lead-acid batteries.
2. Indicate whether each of the following statements is true or false. If the statement is false, rewrite the statement correctly.
- The longer the hydrocarbon chain in an organic compound, the more likely it is to be a solid at room temperature.
 - The main elements used in fertilisers are nitrogen, phosphorus and potassium.
 - A soap molecule is composed of an alcohol molecule and three fatty acids.
 - During the industrial preparation of chlorine and sodium hydroxide, chemical energy is converted to electrical energy.
3. For each of the following questions, choose the one correct answer from the list provided.
- The sequence of processes that best describes the conversion of coal to liquid fuel is:
 - coal \rightarrow gas purification \rightarrow SAS reactor \rightarrow liquid hydrocarbon
 - coal \rightarrow autothermal reactor \rightarrow Sasol slurry phase F-T reactor \rightarrow liquid hydrocarbon
 - coal \rightarrow coal purification \rightarrow synthesis gas \rightarrow oil
 - coal \rightarrow coal gasification \rightarrow gas purification \rightarrow SAS reactor \rightarrow liquid hydrocarbons
 - The half-reaction that takes place at the cathode of a mercury cell in the chloralkali industry is:
 - $2\text{Cl}^- \rightarrow \text{Cl}_2 + 2\text{e}^-$
 - $2\text{Na}^+ + 2\text{e}^- \rightarrow 2\text{Na}$
 - $2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2$
 - $\text{NaCl} + \text{H}_2\text{O} \rightarrow \text{NaOH} + \text{HCl}$
 - In a zinc-carbon dry cell...
 - the electrolyte is manganese (IV) oxide
 - zinc is oxidised to produce electrons
 - zinc is reduced to produce electrons
 - manganese (IV) dioxide acts as a reducing agent

4. Chloralkali manufacturing process

The chloralkali (also called 'chlorine-caustic') industry is one of the largest electrochemical technologies in the world. Chlorine is produced using three types of electrolytic cells. The simplified diagram below shows a membrane cell.

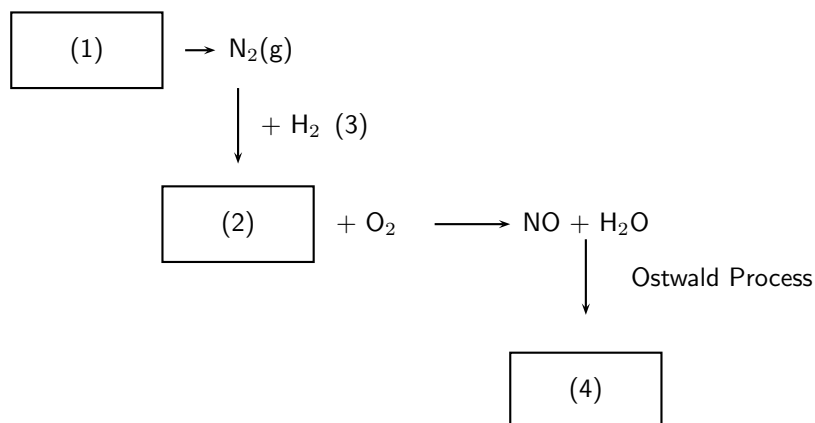


- Give two reasons why the membrane cell is the preferred cell for the preparation of chlorine.
- Why do you think it is advisable to use inert electrodes in this process?
- Write down the equation for the half-reaction taking place at electrode M.
- Which gas is chlorine gas? Write down only Gas A or Gas B.

(e) Briefly explain how sodium hydroxide forms in this cell.

(DoE Exemplar Paper 2,2007)

5. The production of nitric acid is very important in the manufacture of fertilisers. Look at the diagram below, which shows part of the fertiliser production process, and then answer the questions that follow.



- (a) Name the process at (1).
 (b) Name the gas at (2).
 (c) Name the process at (3) that produces gas (2).
 (d) Name the product at (4).
 (e) Name two fertilisers that can be produced from nitric acid.
6. A lead-acid battery has a number of different components. Match the description in Column A with the correct word or phrase in Column B. All the descriptions in Column A relate to lead-acid batteries.

Column A

The electrode metal
 Electrolyte
 A product of the overall cell reaction
 An oxidising agent in the cathode half-reaction
 Type of cells in a lead-acid battery

Column B

Lead sulphate
 Mercury
 Electrolytic
 Lead
 Sulfuric acid
 Ammonium chloride
 Lead oxide
 Galvanic

Part II

Physics

Chapter 6

Motion in Two Dimensions - Grade 12

6.1 Introduction

In Grade 10, we studied motion in one dimension and briefly looked at vertical motion. In this chapter we will discuss vertical motion and also look at motion in two dimensions. In Grade 11, we studied the conservation of momentum and looked at applications in one dimension. In this chapter we will look at momentum in two dimensions.

6.2 Vertical Projectile Motion

In Grade 10, we studied the motion of objects in free fall and we saw that an object in free fall falls with gravitational acceleration g . Now we can consider the motion of objects that are thrown upwards and then fall back to the Earth. We call this *projectile motion* and we will only consider the situation where the object is thrown straight upwards and then falls straight downwards - this means that there is no horizontal displacement of the object, only a vertical displacement.

6.2.1 Motion in a Gravitational Field

When an object is in the earth's gravitational field, it always accelerates downwards, towards the centre of the earth, with a constant acceleration g , no matter whether the object is moving upwards or downwards. This is shown in Figure 6.1.



Important: Projectiles moving upwards or downwards in the earth's gravitational field always accelerate downwards with a constant acceleration g . Note: acceleration means that the velocity *changes*; it either becomes greater *or* smaller.

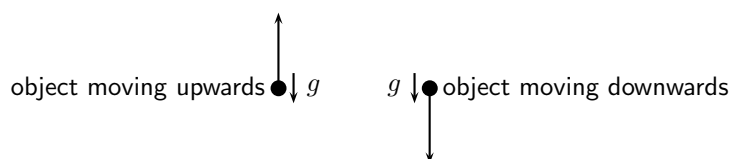


Figure 6.1: Objects moving upwards or downwards, always accelerate downwards.

This means that if an object is moving upwards, its velocity decreases until it stops ($v_f = 0 \text{ m}\cdot\text{s}^{-1}$). This is the maximum height that the object reaches, because after this, the object starts to fall.



Important: Projectiles have zero velocity at their greatest height.

Consider an object thrown upwards from a vertical height h_0 . We have seen that the object will travel upwards with decreasing velocity until it stops, at which point it starts falling. The time that it takes for the object to fall down to height h_0 is the same as the time taken for the object to reach its maximum height from height h_0 .

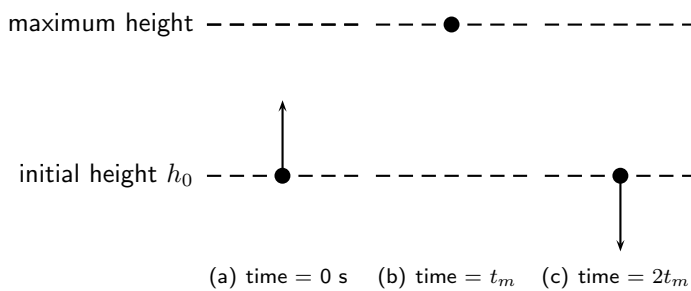


Figure 6.2: (a) An object is thrown upwards from height h_0 . (b) After time t_m , the object reaches its maximum height, and starts to fall. (c) After a time $2t_m$ the object returns to height h_0 .



Important: Projectiles take the same time to go from the point of launch to the greatest height as the time they take to fall back to the point of launch.

6.2.2 Equations of Motion

The equations of motion that were used in Chapter ?? to describe free fall can be used for projectile motion. These equations are the same as those equations that were derived in Chapter ??, but with acceleration from gravity: $a = g$. We use $g = 9,8 \text{ m} \cdot \text{s}^{-2}$ for our calculations.

Remember that when you use these equations, you are dealing with vectors which have magnitude *and* direction. Therefore, you need to decide which direction will be the positive direction so that your vectors have the correct signs.

$$\begin{aligned}
 v_i &= \text{initial velocity (m} \cdot \text{s}^{-1}) \text{ at } t = 0 \text{ s} \\
 v_f &= \text{final velocity (m} \cdot \text{s}^{-1}) \text{ at time } t \\
 \Delta x &= \text{vertical displacement (m)} \\
 t &= \text{time (s)} \\
 \Delta t &= \text{time interval (s)} \\
 g &= \text{acceleration due to gravity (m} \cdot \text{s}^{-2})
 \end{aligned}$$

$$v_f = v_i + gt \quad (6.1)$$

$$\Delta x = \frac{(v_i + v_f)t}{2} \quad (6.2)$$

$$\Delta x = v_i t + \frac{1}{2}gt^2 \quad (6.3)$$

$$v_f^2 = v_i^2 + 2g\Delta x \quad (6.4)$$



Worked Example 29: Projectile motion

Question: A ball is thrown upwards with an initial velocity of $10 \text{ m} \cdot \text{s}^{-1}$.

1. Determine the maximum height reached above the thrower's hand.
2. Determine the time it takes the ball to reach its maximum height.

Answer**Step 1 : Identify what is required and what is given**

We are required to determine the maximum height reached by the ball and how long it takes to reach this height. We are given the initial velocity $v_i = 10 \text{ m}\cdot\text{s}^{-1}$ and the acceleration due to gravity $g = 9,8 \text{ m}\cdot\text{s}^{-2}$.

Step 2 : Determine how to approach the problem

Choose down as positive. We know that at the maximum height the velocity of the ball is $0 \text{ m}\cdot\text{s}^{-1}$. We therefore have the following:

- $v_i = -10 \text{ m}\cdot\text{s}^{-1}$ (it is negative because we chose downwards as positive)
- $v_f = 0 \text{ m}\cdot\text{s}^{-1}$
- $g = +9,8 \text{ m}\cdot\text{s}^{-2}$

Step 3 : Identify the appropriate equation to determine the height.

We can use:

$$v_f^2 = v_i^2 + 2g\Delta x$$

to solve for the height.

Step 4 : Substitute the values in and find the height.

$$\begin{aligned} v_f^2 &= v_i^2 + 2g\Delta x \\ (0)^2 &= (-10)^2 + (2)(9,8)(\Delta x) \\ -100 &= 19,6\Delta x \\ \Delta x &= -5,102\text{m} \end{aligned}$$

The value for the displacement will be negative because the displacement is upwards and we have chosen downward as positive (and upward as negative). The height will be a positive number, $h = 5,10\text{m}$.

Step 5 : Identify the appropriate equation to determine the time.

We can use:

$$v_f = v_i + gt$$

to solve for the time.

Step 6 : Substitute the values in and find the time.

$$\begin{aligned} v_f &= v_i + gt \\ 0 &= -10 + 9,8t \\ 10 &= 9,8t \\ t &= 1,02\text{s} \end{aligned}$$

Step 7 : Write the final answer.

The ball reaches a maximum height of 5,10 m.

The ball takes 1,02 s to reach the top.



Worked Example 30: Height of a projectile

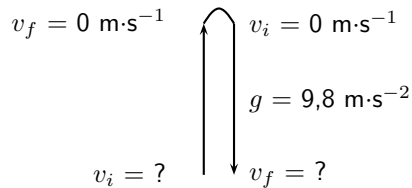
Question: A cricketer hits a cricket ball from the ground so that it goes directly upwards. If the ball takes, 10 s to return to the ground, determine its maximum height.

Answer**Step 1 : Identify what is required and what is given**

We need to find how high the ball goes. We know that it takes 10 seconds to go up and down. We do not know what the initial velocity of the ball (v_i) is.

Step 2 : Determine how to approach the problem

A problem like this one can be looked at as if there are two parts to the motion. The first is the ball going up with an initial velocity and stopping at the top (final velocity is zero). The second motion is the ball falling, its initial velocity is zero and its final velocity is unknown.



Choose down as positive. We know that at the maximum height, the velocity of the ball is $0 \text{ m}\cdot\text{s}^{-1}$. We also know that the ball takes the same time to reach its maximum height as it takes to travel from its maximum height to the ground. This time is half the total time. We therefore know the following for the second motion of the ball going down:

- $t = 5 \text{ s}$, half of the total time
- $v_{top} = v_i = 0 \text{ m}\cdot\text{s}^{-1}$
- $g = 9,8 \text{ m}\cdot\text{s}^{-2}$
- $\Delta x = ?$

Step 3 : Find an appropriate equation to use

We do not know the final velocity of the ball coming down. We need to choose an equation that does not have v_f in it. We can use the following equation to solve for Δx :

$$\Delta x = v_i t + \frac{1}{2} g t^2$$

Step 4 : Substitute values and find the height.

$$\begin{aligned} \Delta x &= (0)(5) + \frac{1}{2}(9,8)(5)^2 \\ \Delta x &= 0 + 122,5\text{m} \end{aligned}$$

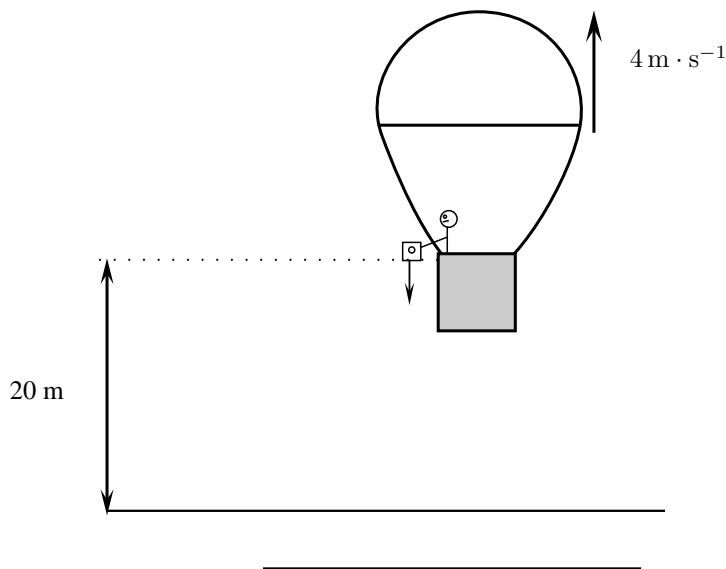
In the second motion, the displacement of the ball is 122,5m downwards. This means that the height was 122,5m, $h=122,5\text{m}$.

Step 5 : Write the final answer

The ball reaches a maximum height of 122,5 m.



- A cricketer hits a cricket ball straight up into the air. The cricket ball has an initial velocity of $20 \text{ m}\cdot\text{s}^{-1}$.
 - What height does the ball reach before it stops to fall back to the ground.
 - How long has the ball been in the air for?
- Zingi throws a tennis ball straight up into the air. It reaches a height of 80 cm.
 - Determine the initial velocity of the tennis ball.
 - How long does the ball take to reach its maximum height?
- A tourist takes a trip in a hot air balloon. The hot air balloon is ascending (moving up) at a velocity of $4 \text{ m}\cdot\text{s}^{-1}$. He accidentally drops his camera over the side of the balloon's basket, at a height of 20 m. Calculate the velocity with which the camera hits the ground.



6.2.3 Graphs of Vertical Projectile Motion

Vertical projectile motion is the same as motion at constant acceleration. In Grade 10 you learned about the graphs for motion at constant acceleration. The graphs for vertical projectile motion are therefore identical to the graphs for motion under constant acceleration.

When we draw the graphs for vertical projectile motion, we consider two main situations: an object moving upwards and an object moving downwards.

If we take the upwards direction as positive then for an object moving upwards we get the graphs shown in Figure 11.



Worked Example 31: Drawing Graphs of Projectile Motion

Question: Stanley is standing on the a balcony 20 m above the ground. Stanley tosses up a rubber ball with an initial velocity of $4,9 \text{ m}\cdot\text{s}^{-1}$. The ball travels upwards and then falls to the ground. Draw graphs of position vs. time, velocity vs. time and acceleration vs. time. Choose upwards as the positive direction.

Answer

Step 1 : Determine what is required

We are required to draw graphs of

- Δx vs. t
- v vs. t
- a vs. t

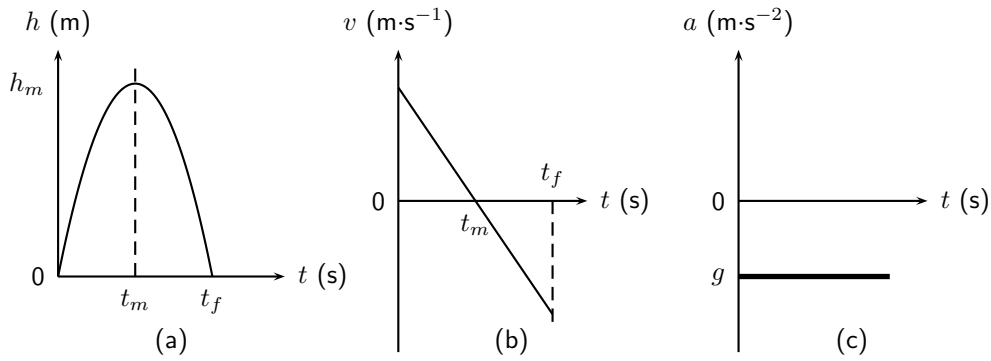


Figure 6.3: Graphs for an object thrown upwards with an initial velocity v_i . The object takes t_m s to reach its maximum height of h_m m after which it falls back to the ground. (a) position vs. time graph (b) velocity vs. time graph (c) acceleration vs. time graph.

v_f

Step 2 : Determine how to approach the problem

There are two parts to the motion of the ball:

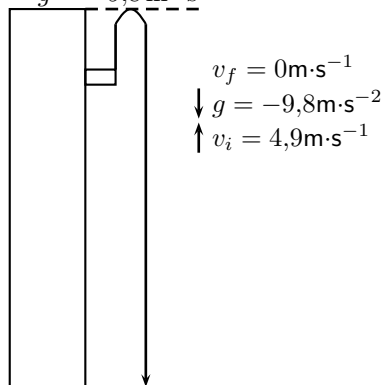
1. ball travelling upwards from the building
2. ball falling to the ground

We examine each of these parts separately. To be able to draw the graphs, we need to determine the time taken and displacement for each of the motions.

Step 3 : Find the height and the time taken for the first motion.

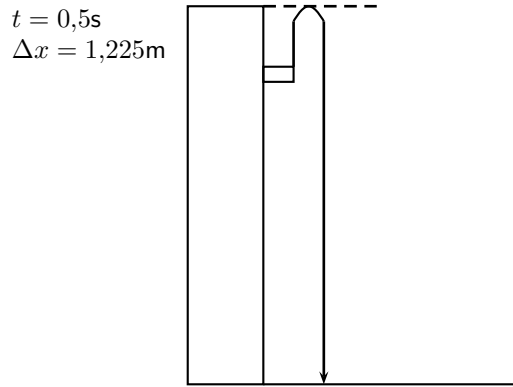
For the first part of the motion we have:

- $v_i = +4,9 \text{ m} \cdot \text{s}^{-1}$
- $v_f = 0 \text{ m} \cdot \text{s}^{-1}$
- $g = -9,8 \text{ m} \cdot \text{s}^{-2}$



Therefore we can use $v_f^2 = v_i^2 + 2g\Delta x$ to solve for the height and $v_f = v_i + gt$ to solve for the time.

$$\begin{aligned}
 v_f^2 &= v_i^2 + 2g\Delta x & v_f &= v_i + gt \\
 (0)^2 &= (4,9)^2 + 2 \times (-9,8) \times \Delta x & 0 &= 4,9 + (-9,8) \times t \\
 19,6\Delta x &= (4,9)^2 & 9,8t &= 4,9 \\
 \Delta x &= 1,225 \text{ m} & t &= 0,5 \text{ s}
 \end{aligned}$$



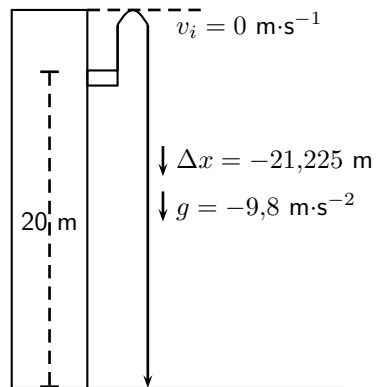
Step 4 : Find the height and the time taken for the second motion.

For the second part of the motion we have:

- $v_i = 0\text{ m} \cdot \text{s}^{-1}$
- $\Delta x = -(20 + 1,225)\text{ m}$
- $g = -9,8\text{ m} \cdot \text{s}^{-2}$

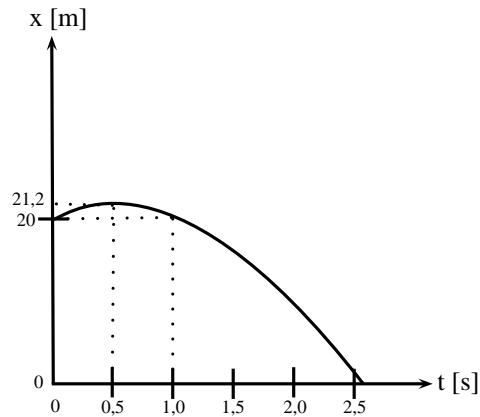
Therefore we can use $\Delta x = v_i t + \frac{1}{2}gt^2$ to solve for the time.

$$\begin{aligned}\Delta x &= v_i t + \frac{1}{2}gt^2 \\ -(20 + 1,225) &= (0) \times t + \frac{1}{2} \times (-9,8) \times t^2 \\ -21,225 &= 0 - 4,9t^2 \\ t^2 &= 4,33163... \\ t &= 2,08125... \text{ s}\end{aligned}$$



Step 5 : Graph of position vs. time

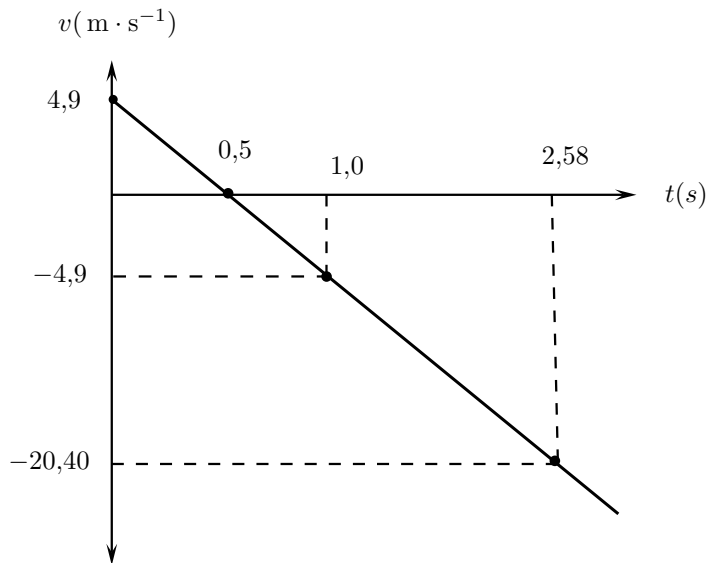
The ball starts from a position of 20 m (at $t = 0\text{ s}$) from the ground and moves upwards until it reaches $(20 + 1,225)\text{ m}$ (at $t = 0,5\text{ s}$). It then falls back to 20 m (at $t = 0,5 + 0,5 = 1,0\text{ s}$) and then falls to the ground, $\Delta x = 0\text{ m}$ at $(t = 0,5 + 2,08 = 2,58\text{ s})$.



Step 6 : Graph of velocity vs. time

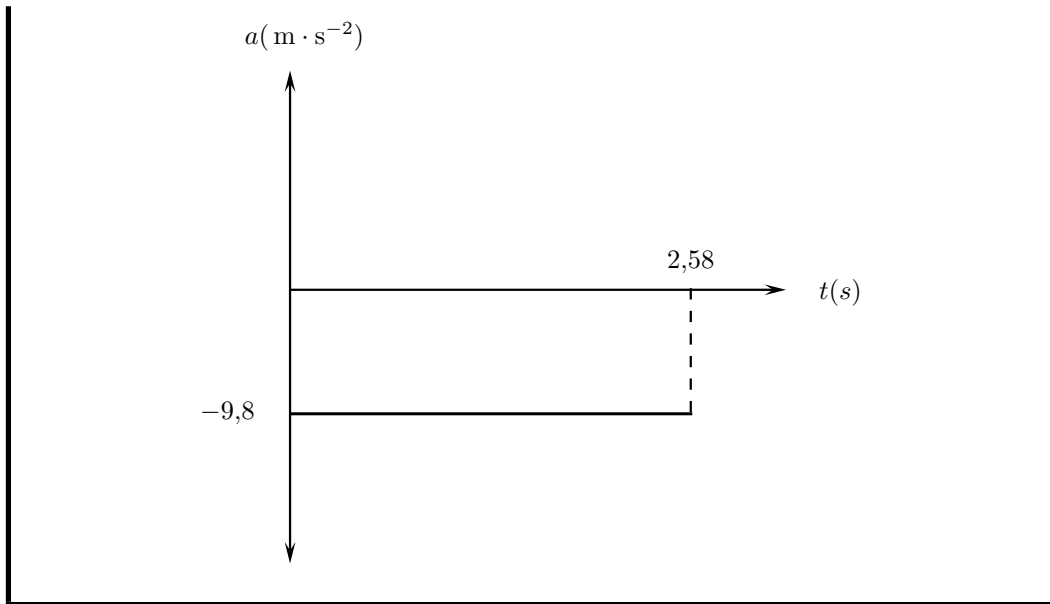
The ball starts off with a velocity of $+4,9 \text{ m}\cdot\text{s}^{-1}$ at $t = 0 \text{ s}$, it then reaches a velocity of $0 \text{ m}\cdot\text{s}^{-1}$ at $t = 0,5 \text{ s}$. It stops and falls back to the Earth. At $t = 1,0 \text{ s}$ (i.e. after a further $0,5 \text{ s}$) it has a velocity of $-4,9 \text{ m}\cdot\text{s}^{-1}$. This is the same as the initial upwards velocity but it is downwards. It carries on at constant acceleration until $t = 2,58 \text{ s}$. In other words, the velocity graph will be a straight line. The final velocity of the ball can be calculated as follows:

$$\begin{aligned} v_f &= v_i + gt \\ &= 0 + (-9,8)(2,08\dots) \\ &= -20,396\dots \text{ m}\cdot\text{s}^{-1} \end{aligned}$$



Step 7 : Graph of a vs t

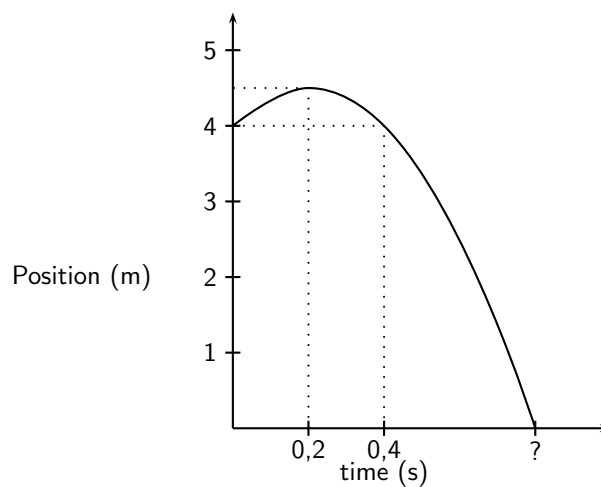
We chose upwards to be positive. The acceleration of the ball is downward. $g = -9,8 \text{ m}\cdot\text{s}^{-2}$. Because the acceleration is constant throughout the motion, the graph looks like this:



Worked Example 32: Analysing Graphs of Projectile Motion

Question: The graph below (not drawn to scale) shows the motion of tennis ball that was thrown vertically upwards from an open window some distance from the ground. It takes the ball $0,2$ s to reach its highest point before falling back to the ground. Study the graph given and calculate

1. how high the window is above the ground.
2. the time it takes the ball to reach the maximum height.
3. the initial velocity of the ball.
4. the maximum height that the ball reaches.
5. the final velocity of the ball when it reaches the ground.



Answer

Step 1 : Find the height of the window.

The initial position of the ball will tell us how high the window is. From the y-axis on the graph we can see that the ball is 4 m from the ground. The window is therefore 4 m above the ground.

Step 2 : Find the time taken to reach the maximum height.

The maximum height is where the position-time graph show the maximum position - the top of the curve. This is when $t = 0,2$ s.
It takes the ball 0,2s to reach the maximum height.

Step 3 : Find the initial velocity (v_i) of the ball.

To find the initial velocity we only look at the first part of the motion of the ball. That is from when the ball is released until it reaches its maximum height. We have the following for this: In this case, let's choose upwards as positive.

$$\begin{aligned}t &= 0,2 \text{ s} \\g &= -9,8 \text{ m} \cdot \text{s}^{-2} \\v_f &= 0 \text{ m} \cdot \text{s}^{-1} \text{ (because the ball stops)}\end{aligned}$$

To calculate the initial velocity of the ball (v_i), we use:

$$\begin{aligned}v_f &= v_i + gt \\0 &= v_i + (-9,8)(0,2) \\v_i &= 1,96 \text{ m} \cdot \text{s}^{-1}\end{aligned}$$

The initial velocity of the ball is $1,96 \text{ m} \cdot \text{s}^{-1}$ upwards.

Step 4 : Find the maximum height (Δx) of the ball.

To find the maximum height we look at the initial motion of the ball. We have the following:

$$\begin{aligned}t &= 0,2 \text{ s} \\g &= -9,8 \text{ m} \cdot \text{s}^{-2} \\v_f &= 0 \text{ m} \cdot \text{s}^{-1} \text{ (because the ball stops)} \\v_i &= +1,96 \text{ m} \cdot \text{s}^{-1} \text{ (calculated above)}\end{aligned}$$

To calculate the displacement from the window to the maximum height (Δx) we use:

$$\begin{aligned}\Delta x &= v_i t + \frac{1}{2} g t^2 \\ \Delta x &= (1,96)(0,2) + \frac{1}{2}(-9,8)(0,2)^2 \\ \Delta x &= 0,196 \text{ m}\end{aligned}$$

The maximum height of the ball is $(4 + 0,196) = 4,196$ m above the ground.

Step 5 : Find the final velocity (v_f) of the ball.

To find the final velocity of the ball we look at the second part of the motion. For this we have:

$$\begin{aligned}\Delta x &= -4,196 \text{ m} \text{ (because upwards is positive)} \\g &= -9,8 \text{ m} \cdot \text{s}^{-2} \\v_i &= 0 \text{ m} \cdot \text{s}^{-1}\end{aligned}$$

We can use $(v_f)^2 = (v_i)^2 + 2g\Delta x$ to calculate the final velocity of the ball.

$$\begin{aligned}(v_f)^2 &= (v_i)^2 + 2g\Delta x \\(v_f)^2 &= (0)^2 + 2(-9,8)(-4,196) \\(v_f)^2 &= 82,2416 \\v_f &= 9,0687... \text{ m} \cdot \text{s}^{-1}\end{aligned}$$

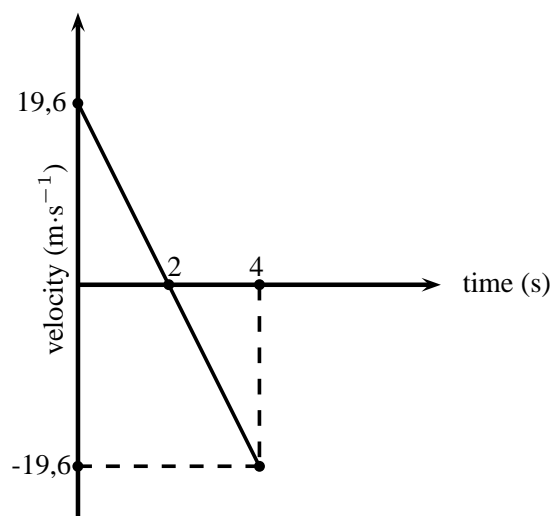
The final velocity of the ball is $9,07 \text{ m} \cdot \text{s}^{-1}$ downwards.



Worked Example 33: Describing Projectile Motion

Question: A cricketer hits a cricket ball from the ground and the following graph of velocity vs. time was drawn. Upwards was taken as positive. Study the graph and follow the instructions below:

1. Describe the motion of the ball according to the graph.
2. Draw a sketch graph of the corresponding displacement-time graph. Label the axes.
3. Draw a sketch graph of the corresponding acceleration-time graph. Label the axes.



Answer

Step 1 : Describe the motion of the ball.

We need to study the velocity-time graph to answer this question. We will break the motion of the ball up into two time zones: $t = 0 \text{ s}$ to $t = 2 \text{ s}$ and $t = 2 \text{ s}$ to $t = 4 \text{ s}$.

From $t = 0 \text{ s}$ to $t = 2 \text{ s}$ the following happens:

The ball starts to move at an initial velocity of $19,6 \text{ m} \cdot \text{s}^{-1}$ and decreases its velocity to $0 \text{ m} \cdot \text{s}^{-1}$ at $t = 2 \text{ s}$. At $t = 2 \text{ s}$ the velocity of the ball is $0 \text{ m} \cdot \text{s}^{-1}$ and therefore it stops.

From $t = 2 \text{ s}$ to $t = 4 \text{ s}$ the following happens:

The ball moves from a velocity of $0 \text{ m} \cdot \text{s}^{-1}$ to $19,6 \text{ m} \cdot \text{s}^{-1}$ in the opposite direction to the original motion.

If we assume that the ball is hit straight up in the air (and we take upwards

as positive), it reaches its maximum height at $t = 2$ s, stops, turns around and falls back to the Earth to reach the ground at $t = 4$ s.

Step 2 : Draw the displacement-time graph.

To draw this graph, we need to determine the displacements at $t = 2$ s and $t = 4$ s.

At $t = 2$ s:

The displacement is equal to the area under the graph:

Area under graph = Area of triangle

$$\text{Area} = \frac{1}{2}bh$$

$$\text{Area} = \frac{1}{2} \times 2 \times 19,6$$

$$\text{Displacement} = 19,6 \text{ m}$$

At $t = 4$ s:

The displacement is equal to the area under the whole graph (top and bottom). Remember that an area under the time line must be subtracted:

Area under graph = Area of triangle 1 + Area of triangle 2

$$\text{Area} = \frac{1}{2}bh + \frac{1}{2}bh$$

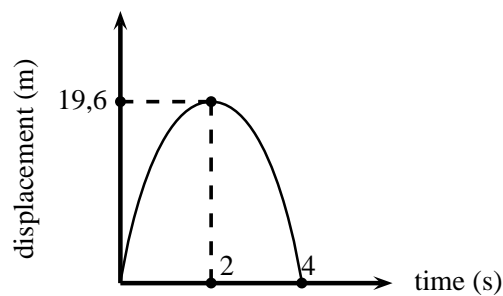
$$\text{Area} = \left(\frac{1}{2} \times 2 \times 19,6\right) + \left(\frac{1}{2} \times 2 \times (-19,6)\right)$$

$$\text{Area} = 19,6 - 19,6$$

$$\text{Displacement} = 0 \text{ m}$$

The displacement-time graph for motion at constant acceleration is a curve.

The graph will look like this:



Step 3 : Draw the acceleration-time graph.

To draw the acceleration vs. time graph, we need to know what the acceleration is. The velocity-time graph is a straight line which means that the acceleration is constant. The gradient of the line will give the acceleration. The line has a negative slope (goes down towards the left) which means that the acceleration has a negative value.

Calculate the gradient of the line:

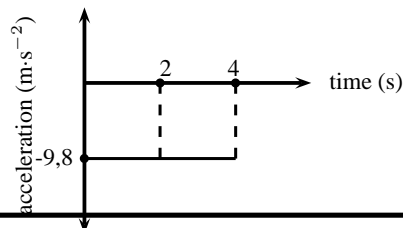
$$\text{gradient} = \frac{\Delta v}{t}$$

$$\text{gradient} = \frac{0 - 19,6}{2 - 0}$$

$$\text{gradient} = \frac{-19,6}{2}$$

$$\text{gradient} = -9,8$$

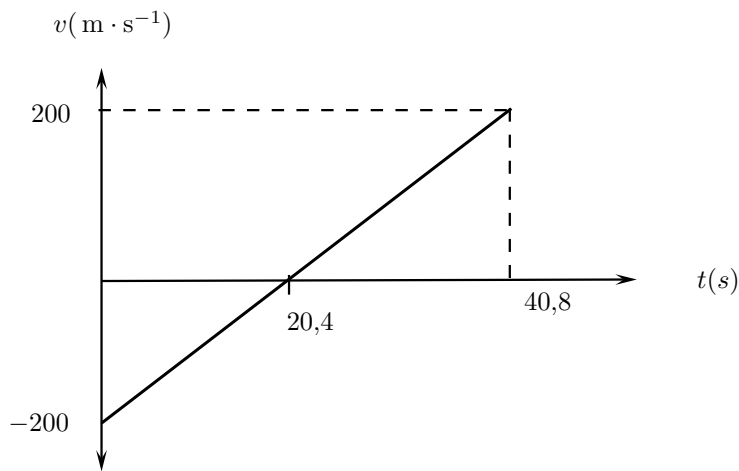
$$\text{acceleration} = 9,8 \text{ m}\cdot\text{s}^{-2} \text{ downwards}$$





Exercise: Graphs of Vertical Projectile Motion

1. Amanda throws a tennisball from a height of 1,5m straight up into the air and then lets it fall to the ground. Draw graphs of Δx vs t ; v vs t and a vs t for the motion of the ball. The initial velocity of the tennisball is $2 \text{ m} \cdot \text{s}^{-1}$. Choose upwards as positive.
2. A bullet is shot straight upwards from a gun. The following graph is drawn. Downwards was chosen as positive
 - a Describe the motion of the bullet.
 - b Draw a displacement - time graph.
 - c Draw an acceleration - time graph.



6.3 Conservation of Momentum in Two Dimensions

We have seen in Grade 11 that the momentum of a system is conserved when there are no external forces acting on the system. Conversely, an external force causes a change in momentum Δp , with the impulse delivered by the force, F acting for a time Δt given by:

$$\Delta p = F \cdot \Delta t$$

The same principles that were studied in applying the conservation of momentum to problems in one dimension, can be applied to solving problems in two dimensions.

The calculation of momentum is the same in two dimensions as in one dimension. The calculation of momentum in two dimensions is broken down into determining the x and y components of momentum and applying the conservation of momentum to each set of components.

Consider two objects moving towards each other as shown in Figure 6.4. We analyse this situation by calculating the x and y components of the momentum of each object.

Before the collision

Total momentum:

$$p_{i1} = m_1 v_{i1}$$

$$p_{i2} = m_2 v_{i2}$$

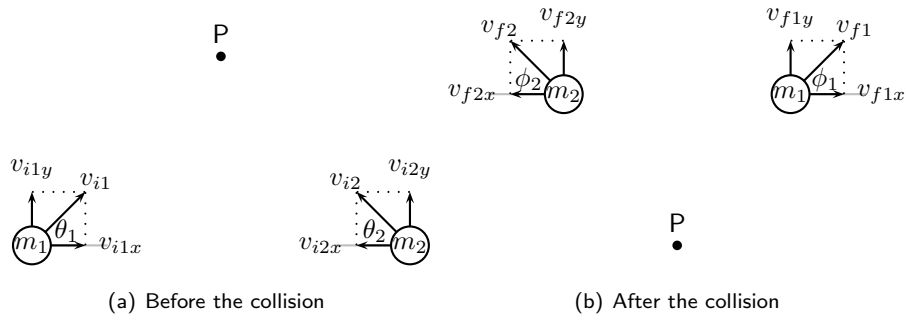


Figure 6.4: Two balls collide at point P.

x-component of momentum:

$$p_{i1x} = m_1 v_{i1x} = m_1 v_{i1} \cos \theta_1$$

$$p_{i2x} = m_2 v_{i2x} = m_2 v_{i2} \cos \theta_2$$

y-component of momentum:

$$p_{i1y} = m_1 v_{i1y} = m_1 v_{i1} \sin \theta_1$$

$$p_{i2y} = m_2 v_{i2y} = m_2 v_{i2} \sin \theta_2$$

After the collision

Total momentum:

$$p_{f1} = m_1 v_{f1}$$

$$p_{f2} = m_2 v_{f2}$$

x-component of momentum:

$$p_{f1x} = m_1 v_{f1x} = m_1 v_{f1} \cos \phi_1$$

$$p_{f2x} = m_2 v_{f2x} = m_2 v_{f2} \cos \phi_2$$

y-component of momentum:

$$p_{f1y} = m_1 v_{f1y} = m_1 v_{f1} \sin \phi_1$$

$$p_{f2y} = m_2 v_{f2y} = m_2 v_{f2} \sin \phi_2$$

Conservation of momentum

The initial momentum is equal to the final momentum:

$$p_i = p_f$$

$$p_i = p_{i1} + p_{i2}$$

$$p_f = p_{f1} + p_{f2}$$

This forms the basis of analysing momentum conservation problems in two dimensions.



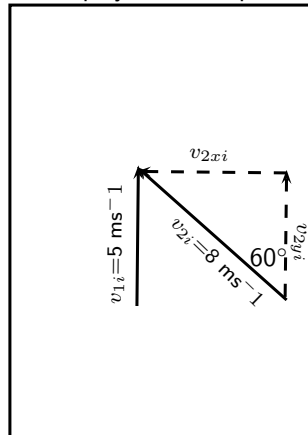
Worked Example 34: 2D Conservation of Momentum

Question: In a rugby game, Player 1 is running with the ball at $5 \text{ m}\cdot\text{s}^{-1}$ straight down the field parallel to the edge of the field. Player 2 runs at $6 \text{ m}\cdot\text{s}^{-1}$ an angle of 60° to the edge of the field and tackles Player 1. In the tackle, Player 2 stops completely while Player 1 bounces off Player 2. Calculate the velocity (magnitude and direction) at which Player 1 bounces off Player 2. Both the players have a mass of 90 kg .

Answer

Step 1 : Identify what is required and what is given

The first step is to draw the picture to work out what the situation is. Mark the initial velocities of both players in the picture.



We also know that $m_1 = m_2 = 90 \text{ kg}$ and $v_{f2} = 0 \text{ m}\cdot\text{s}^{-1}$.

We need to find the final velocity and angle at which Player 1 bounces off Player 2.

Step 2 : Use conservation of momentum to solve the problem. First find the initial total momentum:

Total initial momentum = Total final momentum. But we have a two-dimensional problem, and we need to break up the initial momentum into x and y components.

$$p_{ix} = p_{fx}$$

$$p_{iy} = p_{fy}$$

For Player 1:

$$p_{ix1} = m_1 v_{i1x} = 90 \times 0 = 0$$

$$p_{iy1} = m_1 v_{i1y} = 90 \times 5$$

For Player 2:

$$p_{ix2} = m_2 v_{i2x} = 90 \times 8 \times \sin 60^\circ$$

$$p_{iy2} = m_2 v_{i2y} = 90 \times 8 \times \cos 60^\circ$$

Step 3 : Now write down what we know about the final momentum:

For Player 1:

$$p_{fx1} = m_1 v_{fx1} = 90 \times v_{fx1}$$

$$p_{fy1} = m_1 v_{fy1} = 90 \times v_{fy1}$$

For Player 2:

$$p_{fx2} = m_2 v_{fx2} = 90 \times 0 = 0$$

$$p_{fy2} = m_2 v_{fy2} = 90 \times 0 = 0$$

Step 4 : Use conservation of momentum:

The initial total momentum in the x direction is equal to the final total momentum in the x direction.

The initial total momentum in the y direction is equal to the final total momentum in the y direction.



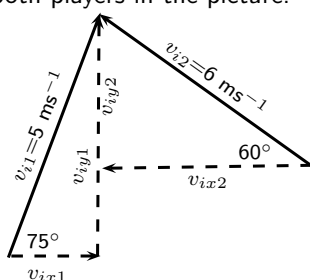
Worked Example 35: 2D Conservation of Momentum: II

Question: In a soccer game, Player 1 is running with the ball at $5 \text{ m}\cdot\text{s}^{-1}$ across the pitch at an angle of 75° from the horizontal. Player 2 runs towards Player 1 at $6 \text{ m}\cdot\text{s}^{-1}$ an angle of 60° to the horizontal and tackles Player 1. In the tackle, the two players bounce off each other. Player 2 moves off with a velocity in the opposite x -direction of $0.3 \text{ m}\cdot\text{s}^{-1}$ and a velocity in the y -direction of $6 \text{ m}\cdot\text{s}^{-1}$. Both the players have a mass of 80 kg . What is the final total velocity of Player 1?

Answer

Step 1 : Identify what is required and what is given

The first step is to draw the picture to work out what the situation is. Mark the initial velocities of both players in the picture.



We need to define a reference frame: For y , choose the direction they are both running in as positive. For x , the direction player 2 is running in is positive.

We also know that $m_1 = m_2 = 80 \text{ kg}$. And $v_{fx2} = -0.3 \text{ m}\cdot\text{s}^{-1}$ and $v_{fy2} = 6 \text{ m}\cdot\text{s}^{-1}$.

We need to find the final velocity and angle at which Player 1 bounces off Player 2.

Step 2 : Use conservation of momentum to solve the problem. First find the initial total momentum:

Total initial momentum = Total final momentum. But we have a two-dimensional problem, and we need to break up the initial momentum into x and y components. Remember that momentum is a vector and has direction which we will indicate with a '+' or '-' sign.

$$p_{ix} = p_{fx}$$

$$p_{iy} = p_{fy}$$

For Player 1:

$$p_{ix1} = m_1 v_{i1x} = 80 \times (-5) \times \cos 75^\circ$$

$$p_{iy1} = m_1 v_{i1y} = 80 \times 5 \times \sin 75^\circ$$

For Player 2:

$$p_{ix2} = m_2 v_{i2x} = 80 \times 6 \times \cos 60^\circ$$

$$p_{iy2} = m_2 v_{i2y} = 80 \times 6 \times \sin 60^\circ$$

Step 3 : Now write down what we know about the final momentum:

For Player 1:

$$p_{fx1} = m_1 v_{fx1} = 80 \times v_{fx1}$$

$$p_{fy1} = m_1 v_{fy1} = 80 \times v_{fy1}$$

For Player 2:

$$p_{fx2} = m_2 v_{fx2} = 80 \times (-0.3)$$

$$p_{fy2} = m_2 v_{fy2} = 80 \times 6$$

Step 4 : Use conservation of momentum:

The initial total momentum in the x direction is equal to the final total momentum in the x direction.

The initial total momentum in the y direction is equal to the final total momentum in the y direction.

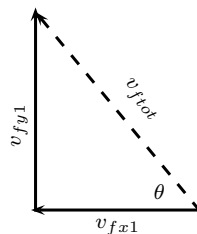
If we find the final x and y components, then we can find the final *total* momentum.

$$\begin{aligned} p_{ix1} + p_{ix2} &= p_{fx1} + p_{fx2} \\ -80 \times 5 \cos 75^\circ + 80 \times 6 \times \cos 60^\circ &= 80 \times v_{fx1} + 80 \times (-0.3) \\ v_{fx1} &= \frac{-80 \times 5 \cos 75^\circ + 80 \times 6 \times \cos 60^\circ - 80 \times (-0.3)}{80} \\ v_{fx1} &= 2.0 \text{ ms}^{-1} \end{aligned}$$

$$\begin{aligned} p_{iy1} + p_{iy2} &= p_{fy1} + p_{fy2} \\ 80 \times 5 \times \sin 75^\circ + 80 \times 6 \times \sin 60^\circ &= 80 \times v_{fy1} + 80 \times 6 \\ v_{fy1} &= \frac{80 \times 5 \sin 75^\circ + 80 \times 6 \times \sin 60^\circ - 80 \times 6}{80} \\ v_{fy1} &= 4.0 \text{ ms}^{-1} \end{aligned}$$

Step 5 : Using the x and y components, calculate the final total v

Use Pythagoras's theorem to find the total final velocity:



$$\begin{aligned} v_{ftot} &= \sqrt{v_{fx1}^2 + v_{fy1}^2} \\ &= \sqrt{2^2 + 4^2} \\ &= 4.5 \end{aligned}$$

Calculate the angle θ to find the direction of Player 1's final velocity:

$$\begin{aligned} \tan \theta &= \frac{v_{fy1}}{v_{fx1}} \\ \theta &= 63.4^\circ \end{aligned}$$

Therefore Player 1 bounces off Player 2 with a final velocity of $4.5 \text{ m} \cdot \text{s}^{-1}$ at an angle of 63.4° from the horizontal.

6.4 Types of Collisions

Two types of collisions are of interest:

- elastic collisions
- inelastic collisions

In both types of collision, total momentum is *always* conserved. Kinetic energy is conserved for elastic collisions, but not for inelastic collisions.

6.4.1 Elastic Collisions



Definition: Elastic Collisions

An elastic collision is a collision where total momentum and total kinetic energy are both conserved.

This means that in an elastic collision the total momentum *and* the total kinetic energy before the collision is the same as after the collision. For these kinds of collisions, the kinetic energy is not changed into another type of energy.

Before the Collision

Figure 6.5 shows two balls rolling toward each other, about to collide:



Figure 6.5: Two balls before they collide.

Before the balls collide, the total momentum of the system is equal to all the individual momenta added together. Ball 1 has a momentum which we call p_{i1} and ball 2 has a momentum which we call p_{i2} , it means the total momentum before the collision is:

$$p_i = p_{i1} + p_{i2}$$

We calculate the total kinetic energy of the system in the same way. Ball 1 has a kinetic energy which we call KE_{i1} and the ball 2 has a kinetic energy which we call KE_{i2} , it means that the total kinetic energy before the collision is:

$$KE_i = KE_{i1} + KE_{i2}$$

After the Collision

Figure 6.6 shows two balls after they have collided:

After the balls collide and bounce off each other, they have new momenta and new kinetic energies. Like before, the total momentum of the system is equal to all the individual momenta added together. Ball 1 now has a momentum which we call p_{f1} and ball 2 now has a momentum which we call p_{f2} , it means the total momentum after the collision is

$$p_f = p_{f1} + p_{f2}$$

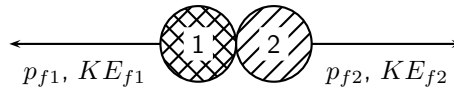


Figure 6.6: Two balls after they collide.

Ball 1 now has a kinetic energy which we call KE_{f1} and ball 2 now has a kinetic energy which we call KE_{f2} , it means that the total kinetic energy after the collision is:

$$KE_f = KE_{f1} + KE_{f2}$$

Since this is an *elastic* collision, the total momentum before the collision equals the total momentum after the collision **and** the total kinetic energy before the collision equals the total kinetic energy after the collision. Therefore:

$$\begin{array}{rcl} \text{Initial} & & \text{Final} \\ p_i & = & p_f \\ p_{i1} + p_{i2} & = & p_{f1} + p_{f2} \\ \text{and} & & \\ KE_i & = & KE_f \\ KE_{i1} + KE_{i2} & = & KE_{f1} + KE_{f2} \end{array}$$



Worked Example 36: An Elastic Collision

Question: Consider a collision between two pool balls. Ball 1 is at rest and ball 2 is moving towards it with a speed of $2 \text{ m}\cdot\text{s}^{-1}$. The mass of each ball is 0.3 kg . After the balls collide *elastically*, ball 2 comes to an immediate stop and ball 1 moves off. What is the final velocity of ball 1?

Answer

Step 1 : Determine how to approach the problem

We are given:

- mass of ball 1, $m_1 = 0.3 \text{ kg}$
- mass of ball 2, $m_2 = 0.3 \text{ kg}$
- initial velocity of ball 1, $v_{i1} = 0 \text{ m}\cdot\text{s}^{-1}$
- initial velocity of ball 2, $v_{i2} = 2 \text{ m}\cdot\text{s}^{-1}$
- final velocity of ball 2, $v_{f2} = 0 \text{ m}\cdot\text{s}^{-1}$
- the collision is elastic

All quantities are in SI units. We are required to determine the final velocity of ball 1, v_{f1} . Since the collision is elastic, we know that

- momentum is conserved, $m_1v_{i1} + m_2v_{i2} = m_1v_{f1} + m_2v_{f2}$
- energy is conserved, $\frac{1}{2}(m_1v_{i1}^2 + m_2v_{i2}^2) = \frac{1}{2}(m_1v_{f1}^2 + m_2v_{f2}^2)$

Step 2 : Choose a frame of reference

Choose to the right as positive.

Step 3 : Draw a rough sketch of the situation



Step 4 : Solve the problem

Momentum is conserved. Therefore:

$$\begin{aligned} p_i &= p_f \\ m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\ (0,3)(0) + (0,3)(2) &= (0,3)v_{f1} + 0 \\ v_{f1} &= 2 \text{ m} \cdot \text{s}^{-1} \end{aligned}$$

Step 5 : Write down the final answer

The final velocity of ball 1 is $2 \text{ m} \cdot \text{s}^{-1}$ to the right.

**Worked Example 37: Another Elastic Collision**

Question: Consider two marbles. Marble 1 has mass 50 g and marble 2 has mass 100 g. Edward rolls marble 2 along the ground towards marble 1 in the positive x -direction. Marble 1 is initially at rest and marble 2 has a velocity of $3 \text{ m} \cdot \text{s}^{-1}$ in the positive x -direction. After they collide *elastically*, both marbles are moving. What is the final velocity of each marble?

Answer

Step 1 : Decide how to approach the problem

We are given:

- mass of marble 1, $m_1=50 \text{ g}$
- mass of marble 2, $m_2=100 \text{ g}$
- initial velocity of marble 1, $v_{i1}=0 \text{ m} \cdot \text{s}^{-1}$
- initial velocity of marble 2, $v_{i2}=3 \text{ m} \cdot \text{s}^{-1}$
- the collision is elastic

The masses need to be converted to SI units.

$$\begin{aligned} m_1 &= 0,05 \text{ kg} \\ m_2 &= 0,1 \text{ kg} \end{aligned}$$

We are required to determine the final velocities:

- final velocity of marble 1, v_{f1}
- final velocity of marble 2, v_{f2}

Since the collision is elastic, we know that

- momentum is conserved, $p_i = p_f$.
- energy is conserved, $KE_i = KE_f$.

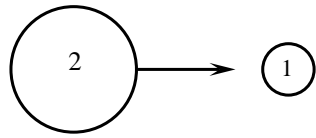
We have two equations and two unknowns (v_1, v_2) so it is a simple case of solving a set of simultaneous equations.

Step 2 : Choose a frame of reference

Choose to the right as positive.

Step 3 : Draw a rough sketch of the situation

Before Collision



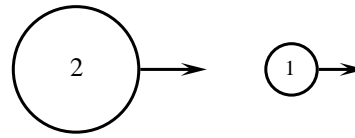
$$m_2 = 100g$$

$$m_1 = 50g$$

$$v_{i2} = 3 \text{ m} \cdot \text{s}^{-1}$$

$$v_{i1} = 0$$

After Collision



$$m_2 = 100g$$

$$m_1 = 50g$$

Step 4 : Solve the problem

Momentum is conserved. Therefore:

$$\begin{aligned} p_i &= p_f \\ p_{i1} + p_{i2} &= p_{f1} + p_{f2} \\ m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\ (0,05)(0) + (0,1)(3) &= (0,05)v_{f1} + (0,1)v_{f2} \\ 0,3 &= 0,05v_{f1} + 0,1v_{f2} \end{aligned} \quad (6.5)$$

Energy is also conserved. Therefore:

$$\begin{aligned} KE_i &= KE_f \\ KE_{i1} + KE_{i2} &= KE_{f1} + KE_{f2} \\ \frac{1}{2}m_1 v_{i1}^2 + \frac{1}{2}m_2 v_{i2}^2 &= \frac{1}{2}m_1 v_{f1}^2 + \frac{1}{2}m_2 v_{f2}^2 \\ \left(\frac{1}{2}\right)(0,05)(0)^2 + \left(\frac{1}{2}\right)(0,1)(3)^2 &= \frac{1}{2}(0,05)(v_{f1})^2 + \left(\frac{1}{2}\right)(0,1)(v_{f2})^2 \\ 0,45 &= 0,025v_{f1}^2 + 0,05v_{f2}^2 \end{aligned} \quad (6.6)$$

Substitute Equation 6.5 into Equation 6.6 and solve for v_{f2} .

$$\begin{aligned} m_2 v_{i2}^2 &= m_1 v_{f1}^2 + m_2 v_{f2}^2 \\ &= m_1 \left(\frac{m_2}{m_1} (v_{i2} - v_{f2}) \right)^2 + m_2 v_{f2}^2 \\ &= m_1 \frac{m_2^2}{m_1^2} (v_{i2} - v_{f2})^2 + m_2 v_{f2}^2 \\ &= \frac{m_2^2}{m_1} (v_{i2} - v_{f2})^2 + m_2 v_{f2}^2 \\ v_{i2}^2 &= \frac{m_2}{m_1} (v_{i2} - v_{f2})^2 + v_{f2}^2 \\ &= \frac{m_2}{m_1} (v_{i2}^2 - 2 \cdot v_{i2} \cdot v_{f2} + v_{f2}^2) + v_{f2}^2 \\ 0 &= \left(\frac{m_2}{m_1} - 1 \right) v_{i2}^2 - 2 \frac{m_2}{m_1} v_{i2} \cdot v_{f2} + \left(\frac{m_2}{m_1} + 1 \right) v_{f2}^2 \\ &= \left(\frac{0,1}{0,05} - 1 \right) (3)^2 - 2 \frac{0,1}{0,05} (3) \cdot v_{f2} + \left(\frac{0,1}{0,05} + 1 \right) v_{f2}^2 \\ &= (2 - 1)(3)^2 - 2 \cdot 2(3) \cdot v_{f2} + (2 + 1)v_{f2}^2 \\ &= 9 - 12v_{f2} + 3v_{f2}^2 \\ &= 3 - 4v_{f2} + v_{f2}^2 \\ &= (v_{f2} - 3)(v_{f2} - 1) \end{aligned}$$

Substituting back into Equation 6.5, we get:

$$\begin{aligned} v_{f1} &= \frac{m_2}{m_1}(v_{i2} - v_{f2}) \\ &= \frac{0.1}{0.05}(3 - 3) \\ &= 0 \text{ m} \cdot \text{s}^{-1} \\ \text{or} \\ v_{f1} &= \frac{m_2}{m_1}(v_{i2} - v_{f2}) \\ &= \frac{0.1}{0.05}(3 - 1) \\ &= 4 \text{ m} \cdot \text{s}^{-1} \end{aligned}$$

But according to the question, marble 1 is moving after the collision, therefore marble 1 moves to the right at $4 \text{ m} \cdot \text{s}^{-1}$. Substituting this value for v_{f1} into Equation 6.5, we can calculate:

$$\begin{aligned} v_{f2} &= \frac{0.3 - 0.05v_{f1}}{0.1} \\ &= \frac{0.3 - 0.05 \times 4}{0.1} \\ &= 1 \text{ m} \cdot \text{s}^{-1} \end{aligned}$$

Therefore marble 2 moves to the right with a velocity of $1 \text{ m} \cdot \text{s}^{-1}$.



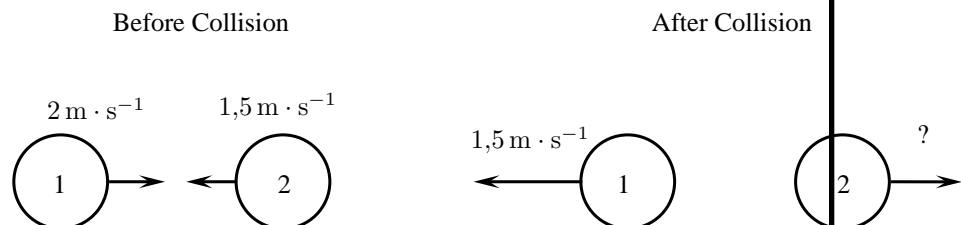
Worked Example 38: Colliding Billiard Balls

Question: Two billiard balls each with a mass of 150 g collide head-on in an elastic collision. Ball 1 was travelling at a speed of $2 \text{ m} \cdot \text{s}^{-1}$ and ball 2 at a speed of $1.5 \text{ m} \cdot \text{s}^{-1}$. After the collision, ball 1 travels away from ball 2 at a velocity of $1.5 \text{ m} \cdot \text{s}^{-1}$.

1. Calculate the velocity of ball 2 after the collision.
2. Prove that the collision was elastic. Show calculations.

Answer

1. **Step 1 : Draw a rough sketch of the situation**



2. **Step 2 : Decide how to approach the problem**

Since momentum is conserved in all kinds of collisions, we can use conservation of momentum to solve for the velocity of ball 2 after the collision.

3. **Step 3 : Solve problem**

$$\begin{aligned}
 p_{before} &= p_{after} \\
 m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\
 \left(\frac{150}{1000}\right)(2) + \left(\frac{150}{1000}\right)(-1,5) &= \left(\frac{150}{1000}\right)(-1,5) + \left(\frac{150}{1000}\right)(v_{f2}) \\
 0,3 - 0,225 &= -0,225 + 0,15v_{f2} \\
 v_{f2} &= 3 \text{ m} \cdot \text{s}^{-1}
 \end{aligned}$$

So after the collision, ball 2 moves with a velocity of $3 \text{ m} \cdot \text{s}^{-1}$.

2. The fact that characterises an elastic collision is that the total kinetic energy of the particles before the collision is the same as the total kinetic energy of the particles after the collision. This means that if we can show that the initial kinetic energy is equal to the final kinetic energy, we have shown that the collision is elastic.

Calculating the initial total kinetic energy:

$$\begin{aligned}
 EK_{before} &= \frac{1}{2}m_1v_{i1}^2 + \frac{1}{2}m_2v_{i2}^2 \\
 &= \left(\frac{1}{2}\right)(0,15)(2)^2 + \left(\frac{1}{2}\right)(0,15)(-1,5)^2 \\
 &= 0.469\dots J
 \end{aligned}$$

Calculating the final total kinetic energy:

$$\begin{aligned}
 EK_{after} &= \frac{1}{2}m_1v_{f1}^2 + \frac{1}{2}m_2v_{f2}^2 \\
 &= \left(\frac{1}{2}\right)(0,15)(-1,5)^2 + \left(\frac{1}{2}\right)(0,15)(2)^2 \\
 &= 0.469\dots J
 \end{aligned}$$

So $EK_{before} = EK_{after}$ and hence the collision is elastic.

6.4.2 Inelastic Collisions



Definition: Inelastic Collisions

An inelastic collision is a collision in which **total momentum** is conserved but **total kinetic energy** is *not* conserved. The kinetic energy is *transformed* into other kinds of energy.

So the total momentum before an inelastic collisions is the same as after the collision. But the total *kinetic* energy before and after the inelastic collision is *different*. Of course this does not mean that total energy has not been conserved, rather the energy has been *transformed* into another type of energy.

As a rule of thumb, inelastic collisions happen when the colliding objects are distorted in some way. Usually they change their shape. The modification of the shape of an object requires energy and this is where the “missing” kinetic energy goes. A classic example of an inelastic collision is a motor car accident. The cars change shape and there is a noticeable change in the kinetic energy of the cars before and after the collision. This energy was used to bend the metal and deform the cars. Another example of an inelastic collision is shown in Figure 6.7.

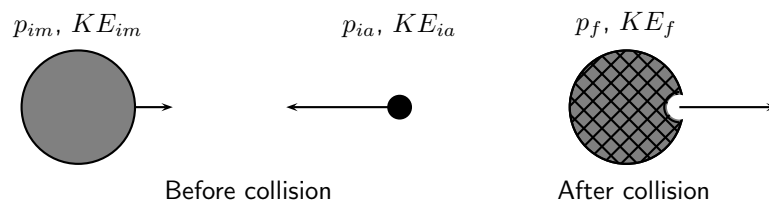


Figure 6.7: Asteroid moving towards the Moon.

An asteroid is moving through space towards the Moon. Before the asteroid crashes into the Moon, the total momentum of the system is:

$$p_i = p_{im} + p_{ia}$$

The total kinetic energy of the system is:

$$KE_i = KE_{im} + KE_{ia}$$

When the asteroid collides **inelastically** with the Moon, its kinetic energy is transformed mostly into heat energy. If this heat energy is large enough, it can cause the asteroid and the area of the Moon's surface that it hits, to melt into liquid rock! From the force of impact of the asteroid, the molten rock flows outwards to form a crater on the Moon.

After the collision, the total momentum of the system will be the same as before. But since this collision is *inelastic*, (and you can see that a change in the shape of objects has taken place!), total kinetic energy is **not** the same as before the collision.

Momentum is conserved:

$$p_i = p_f$$

But the total kinetic energy of the system is not conserved:

$$KE_i \neq KE_f$$

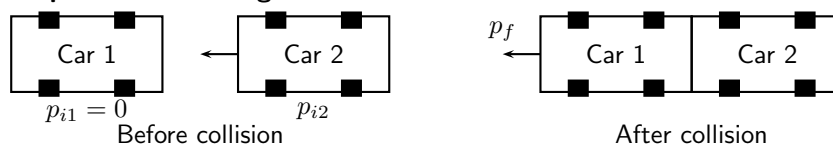


Worked Example 39: An Inelastic Collision

Question: Consider the collision of two cars. Car 1 is at rest and Car 2 is moving at a speed of $2 \text{ m}\cdot\text{s}^{-1}$ to the left. Both cars each have a mass of 500 kg. The cars collide *inelastically* and stick together. What is the resulting velocity of the resulting mass of metal?

Answer

Step 1 : Draw a rough sketch of the situation



Step 2 : Determine how to approach the problem

We are given:

- mass of car 1, $m_1 = 500 \text{ kg}$
- mass of car 2, $m_2 = 500 \text{ kg}$
- initial velocity of car 1, $v_{i1} = 0 \text{ m}\cdot\text{s}^{-1}$
- initial velocity of car 2, $v_{i2} = 2 \text{ m}\cdot\text{s}^{-1}$ to the left
- the collision is inelastic

All quantities are in SI units. We are required to determine the final velocity of the resulting mass, v_f .

Since the collision is inelastic, we know that

- momentum is conserved, $m_1v_{i1} + m_2v_{i2} = m_1v_{f1} + m_2v_{f2}$
- kinetic energy is not conserved

Step 3 : Choose a frame of reference

Choose to the left as positive.

Step 4 : Solve problem

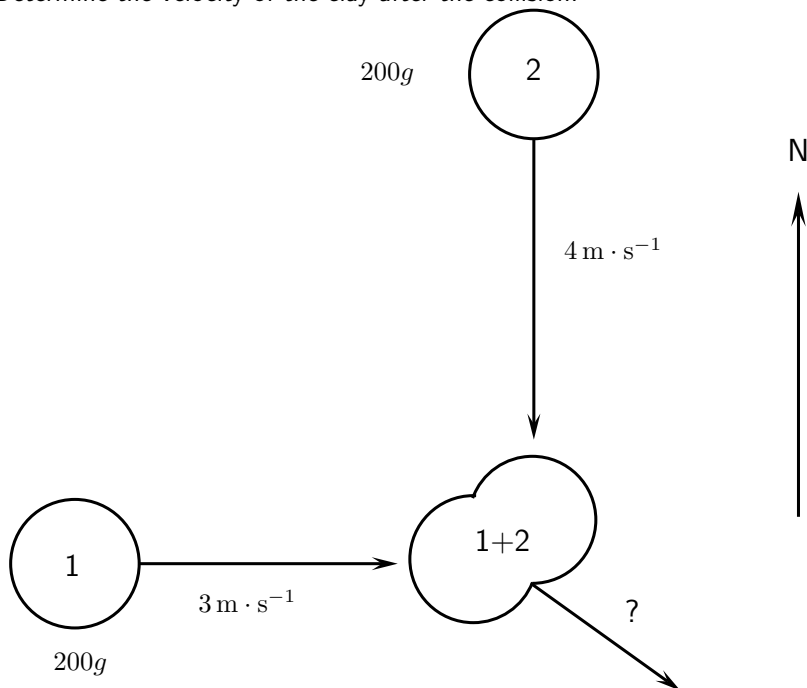
So we must use conservation of momentum to solve this problem.

$$\begin{aligned} p_i &= p_f \\ p_{i1} + p_{i2} &= p_f \\ m_1v_{i1} + m_2v_{i2} &= (m_1 + m_2)v_f \\ (500)(0) + (500)(2) &= (500 + 500)v_f \\ 1000 &= 1000v_f \\ v_f &= 1 \text{ m} \cdot \text{s}^{-1} \end{aligned}$$

Therefore, the final velocity of the resulting mass of cars is $1 \text{ m} \cdot \text{s}^{-1}$ to the left.

**Worked Example 40: Colliding balls of clay**

Question: Two balls of clay, 200g each, are thrown towards each other according to the following diagram. When they collide, they stick together and move off together. All motion is taking place in the horizontal plane. Determine the velocity of the clay after the collision.

**Answer****Step 1 : Analyse the problem**

This is an inelastic collision where momentum is conserved.

The momentum before = the momentum after.

The momentum after can be calculated by drawing a vector diagram.

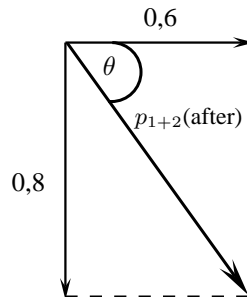
Step 2 : Calculate the momentum before the collision

$$p_1(\text{before}) = m_1 v_{i1} = (0,2)(3) = 0,6 \text{ kg} \cdot \text{m} \cdot \text{s}^{-1} \text{ east}$$

$$p_2(\text{before}) = m_2 v_{i2} = (0,2)(4) = 0,8 \text{ kg} \cdot \text{m} \cdot \text{s}^{-1} \text{ south}$$

Step 3 : Calculate the momentum after the collision.

Here we need to draw a diagram:



$$\begin{aligned} p_{1+2}(\text{after}) &= \sqrt{(0,8)^2 + (0,6)^2} \\ &= 1 \end{aligned}$$

Step 4 : Calculate the final velocity

First we have to find the direction of the final momentum:

$$\begin{aligned} \tan \theta &= \frac{0,8}{0,6} \\ \theta &= 53,13^\circ \end{aligned}$$

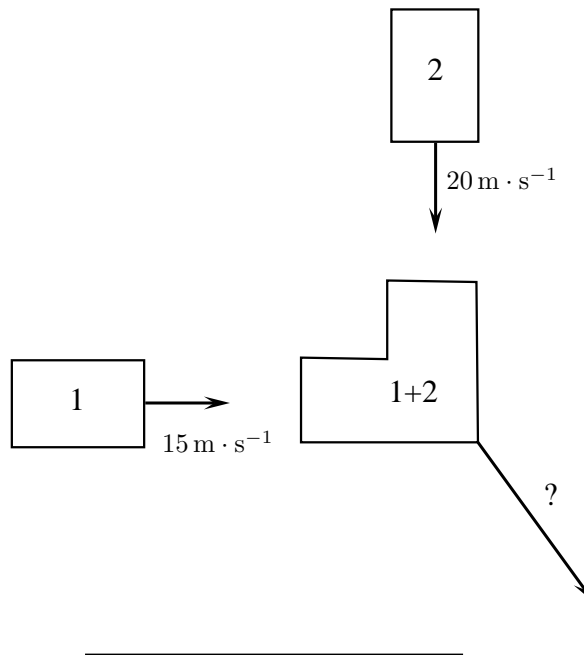
Now we have to find the magnitude of the final velocity:

$$\begin{aligned} p_{1+2} &= m_{1+2} v_f \\ 1 &= (0,2 + 0,2) v_f \\ v_f &= 2,5 \text{ m} \cdot \text{s}^{-1} \quad 53,13^\circ \text{ South of East} \end{aligned}$$



Exercise: Collisions

1. A truck of mass 4500 kg travelling at $20 \text{ m} \cdot \text{s}^{-1}$ hits a car from behind. The car (mass 1000 kg) was travelling at $15 \text{ m} \cdot \text{s}^{-1}$. The two vehicles, now connected carry on moving in the same direction.
 - a Calculate the final velocity of the truck-car combination after the collision.
 - b Determine the kinetic energy of the system before and after the collision.
 - c Explain the difference in your answers for b).
 - d Was this an example of an elastic or inelastic collision? Give reasons for your answer.
2. Two cars of mass 900 kg each collide and stick together at an angle of 90° . Determine the final velocity of the cars if
 - car 1 was travelling at $15 \text{ m} \cdot \text{s}^{-1}$ and
 - car 2 was travelling at $20 \text{ m} \cdot \text{s}^{-1}$.



Extension: Tiny, Violent Collisions

Author: Thomas D. Gutierrez

Tom Gutierrez received his Bachelor of Science and Master degrees in Physics from San Jose State University in his home town of San Jose, California. As a Master's student he helped work on a laser spectrometer at NASA Ames Research Centre. The instrument measured the ratio of different isotopes of carbon in CO_2 gas and could be used for such diverse applications as medical diagnostics and space exploration. Later, he received his PhD in physics from the University of California, Davis where he performed calculations for various reactions in high energy physics collisions. He currently lives in Berkeley, California where he studies proton-proton collisions seen at the STAR experiment at Brookhaven National Laboratory on Long Island, New York.

High Energy Collisions

Take an orange and expand it to the size of the earth. The atoms of the earth-sized orange would themselves be about the size of regular oranges and would fill the entire "earth-orange". Now, take an atom and expand it to the size of a football field. The nucleus of that atom would be about the size of a tiny seed in the middle of the field. From this analogy, you can see that atomic nuclei are very small objects by human standards. They are roughly 10^{-15} meters in diameter – one-hundred thousand times smaller than a typical atom. These nuclei cannot be seen or studied via any conventional means such as the naked eye or microscopes. So how do scientists study the structure of very small objects like atomic nuclei?

The simplest nucleus, that of hydrogen, is called the proton. Faced with the inability to isolate a single proton, open it up, and directly examine what is inside, scientists must resort to a brute-force and somewhat indirect means of exploration: high energy collisions. By colliding protons with other particles (such as other protons or electrons) at very high energies, one hopes to learn about what they are made of and how they work. The American physicist Richard Feynman once compared this process to slamming delicate watches together and figuring out how they work by only examining the broken debris. While this analogy may seem pessimistic, with sufficient mathematical models and experimental precision, considerable information can be extracted from the debris of such high energy subatomic collisions. One can

learn about both the nature of the forces at work and also about the sub-structure of such systems.

The experiments are in the category of “high energy physics” (also known as “subatomic” physics). The primary tool of scientific exploration in these experiments is an extremely violent collision between two very, very small subatomic objects such as nuclei. As a general rule, the higher the energy of the collisions, the more detail of the original system you are able to resolve. These experiments are operated at laboratories such as CERN, SLAC, BNL, and Fermilab, just to name a few. The giant machines that perform the collisions are roughly the size of towns. For example, the RHIC collider at BNL is a ring about 1 km in diameter and can be seen from space. The newest machine currently being built, the LHC at CERN, is a ring 9 km in diameter!

Activity :: Casestudy : Atoms and its Constituents

Questions:

1. What are isotopes? (2)
2. What are atoms made up of? (3)
3. Why do you think protons are used in the experiments and not atoms like carbon? (2)
4. Why do you think it is necessary to find out what atoms are made up of and how they behave during collisions? (2)
5. Two protons (mass $1,67 \times 10^{-27}$ kg) collide and somehow stick together after the collision. If each proton travelled with an initial velocity of $5,00 \times 10^7 \text{ m} \cdot \text{s}^{-1}$ and they collided at an angle of 90° , what is the velocity of the combination after the collision. (9)

6.5 Frames of Reference

6.5.1 Introduction

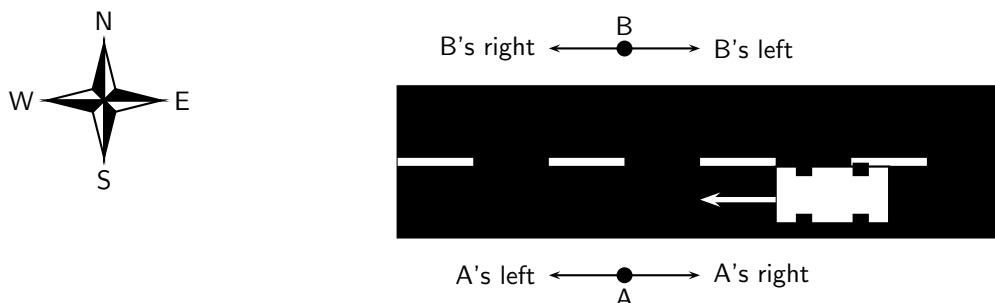


Figure 6.8: Top view of a road with two people standing on opposite sides. A car drives past.

Consider two people standing, facing each other on either side of a road. A car drives past them, heading West. For the person facing South, the car was moving toward the right. However, for the person facing North, the car was moving toward the left. This discrepancy is due to the fact that the two people used two different *frames of reference* from which to investigate this system. If each person were asked in what direction the car were moving, they would give a different answer. The answer would be relative to their frame of reference.

6.5.2 What is a frame of reference?



Definition: Frame of Reference

A *frame of reference* is the point of view from which a system is observed.

In practical terms, a frame of reference is a set of axes (specifying directions) with an origin. An observer can then measure the position and motion of all points in a system, as well as the orientation of objects in the system relative to the frame of reference.

There are two types of reference frames: inertial and non-inertial. An inertial frame of reference travels at a constant velocity, which means that Newton's first law (inertia) holds true. A non-inertial frame of reference, such as a moving car or a rotating carousel, accelerates. Therefore, Newton's first law does not hold true in a non-inertial reference frame, as objects appear to accelerate without the appropriate forces.

6.5.3 Why are frames of reference important?

Frames of reference are important because (as we have seen in the introductory example) the velocity of a car can differ depending on which frame of reference is used.



Extension: Frames of Reference and Special Relativity

Frames of reference are especially important in special relativity, because when a frame of reference is moving at some significant fraction of the speed of light, then the flow of time in that frame does not necessarily apply in another reference frame.

The speed of light is considered to be the only true constant between moving frames of reference.

The next worked example will explain this.

6.5.4 Relative Velocity

The velocity of an object is frame dependent. More specifically, the perceived velocity of an object depends on the velocity of the observer. For example, a person standing on shore would observe the velocity of a boat to be different than a passenger on the boat.

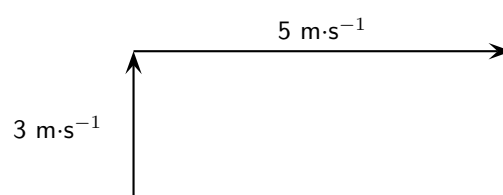


Worked Example 41: Relative Velocity

Question: The speedometer of a motor boat reads $5 \text{ m}\cdot\text{s}^{-1}$. The boat is moving East across a river which has a current traveling $3 \text{ m}\cdot\text{s}^{-1}$ North. What would the velocity of the motor boat be according to an observer on the shore?

Answer

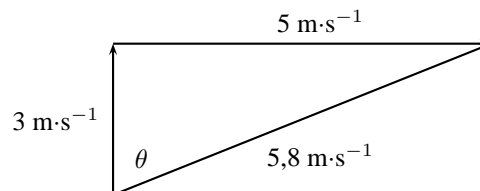
Step 1 : First, draw a diagram showing the velocities involved.



Step 2 : Use the Theorem of Pythagoras to solve for the resultant of the two velocities.

$$\begin{aligned}
 R &= \sqrt{(3)^2 + (5)^2} \\
 &= \sqrt{34} \\
 &= 5,8 \text{ m} \cdot \text{s}^{-1}
 \end{aligned}$$

$$\begin{aligned}
 \tan \theta &= \frac{5}{3} \\
 \theta &= 59,04^\circ
 \end{aligned}$$



The observer on the shore sees the boat moving with a velocity of $5,8 \text{ m} \cdot \text{s}^{-1}$ at $59,04^\circ$ east of north due to the current pushing the boat perpendicular to its velocity. This is contrary to the perspective of a passenger on the boat who perceives the velocity of the boat to be $5 \text{ m} \cdot \text{s}^{-1}$ due East. Both perspectives are correct as long as the frame of the observer is considered.



Worked Example 42: Relative Velocity 2

Question: It takes a man 10 seconds to ride down an escalator. It takes the same man 15 s to walk back up the escalator against its motion. How long will it take the man to walk down the escalator at the same rate he was walking before?

Answer

Step 1 : Identify what is required and what is given

We are required to determine the time taken for a man to walk down an escalator, with its motion.

We are given the time taken for the man to ride down the escalator and the time taken for the man to walk up the escalator, against its motion.

Step 2 : Determine how to approach the problem

Select down as positive and assume that the escalator moves at a velocity v_e . If the distance of the escalator is x_e then:

$$v_e = \frac{x_e}{10 \text{ s}} \quad (6.7)$$

Now, assume that the man walks at a velocity v_m . Then we have that:

$$v_e - v_m = \frac{x_e}{15 \text{ s}} \quad (6.8)$$

We are required to find t in:

$$v_e + v_m = \frac{x_e}{t} \quad (6.9)$$

Step 3 : Solve the problem

We find that we have three equations and three unknowns (v_e , v_m and t).

Add (6.8) to (6.9) to get:

$$2v_e = \frac{x_e}{15 \text{ s}} + \frac{x_e}{t}$$

Substitute from (6.7) to get:

$$2 \frac{x_e}{10 \text{ s}} = \frac{x_e}{15 \text{ s}} + \frac{x_e}{t}$$

Since x_e is not equal to zero we can divide throughout by x_e .

$$\frac{2}{10 \text{ s}} = \frac{1}{15 \text{ s}} + \frac{1}{t}$$

Re-write:

$$\frac{2}{10 \text{ s}} - \frac{1}{15 \text{ s}} = \frac{1}{t}$$

Multiply by t :

$$t \left(\frac{2}{10 \text{ s}} - \frac{1}{15 \text{ s}} \right) = 1$$

Solve for t :

$$t = \frac{1}{\frac{2}{10 \text{ s}} - \frac{1}{15 \text{ s}}}$$

to get:

$$t = \frac{15}{2} \text{ s}$$

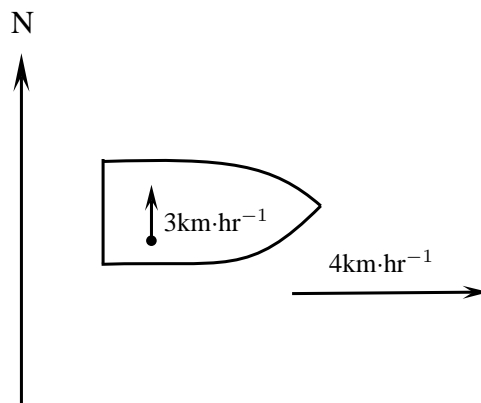
Step 4 : Write the final answer

The man will take $15/2 = 7,5 \text{ s}$



Exercise: Frames of Reference

1. A woman walks north at $3 \text{ km}\cdot\text{hr}^{-1}$ on a boat that is moving east at $4 \text{ km}\cdot\text{hr}^{-1}$. This situation is illustrated in the diagram below.
 - (a) How fast is the woman moving according to her friend who is also on the boat?
 - (b) What is the woman's velocity according to an observer watching from the river bank?



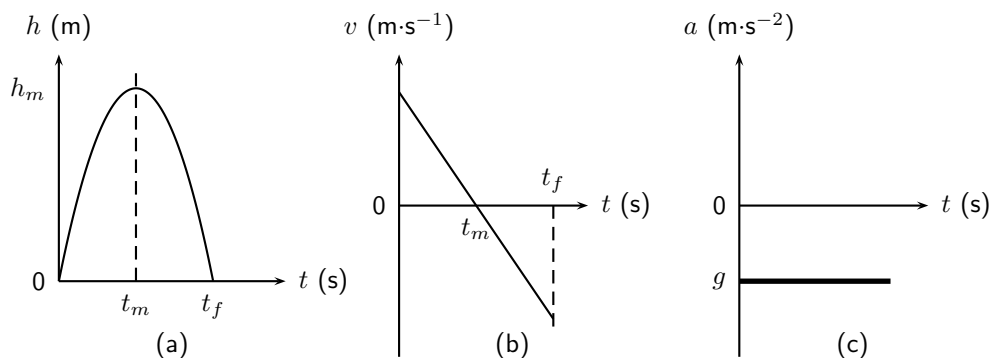
2. A boy is standing inside a train that is moving at $10 \text{ m}\cdot\text{s}^{-1}$ to the left. The boy throws a ball in the air with a velocity of $4 \text{ m}\cdot\text{s}^{-1}$. What is the resultant velocity of the ball
 - (a) according to the boy?
 - (b) according to someone outside the train?

6.6 Summary

1. Projectiles are objects that move through the air.
2. Objects that move up and down (vertical projectiles) on the earth accelerate with a constant acceleration g which is approximately equal to $9,8 \text{ m}\cdot\text{s}^{-2}$ directed downwards towards the centre of the earth.
3. The equations of motion can be used to solve vertical projectile problems.

$$\begin{aligned}v_f &= v_i + gt \\ \Delta x &= \frac{(v_i + v_f)t}{2} \\ \Delta x &= v_i t + \frac{1}{2}gt^2 \\ v_f^2 &= v_i^2 + 2g\Delta x\end{aligned}$$

4. Graphs can be drawn for vertical projectile motion and are similar to the graphs for motion at constant acceleration. If upwards is taken as positive the Δx vs t , v vs t and a vs t graphs for an object being thrown upwards look like this:



5. Momentum is conserved in one and two dimensions.

$$\begin{aligned}p &= mv \\ v_f \Delta p &= m\Delta v \\ \Delta p &= F\Delta t\end{aligned}$$

6. An elastic collision is a collision where both momentum and kinetic energy is conserved.

$$\begin{aligned}p_{\text{before}} &= p_{\text{after}} \\ KE_{\text{before}} &= KE_{\text{after}}\end{aligned}$$

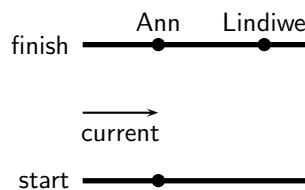
7. An inelastic collision is where momentum is conserved but kinetic energy is not conserved.

$$\begin{aligned}p_{\text{before}} &= p_{\text{after}} \\ KE_{\text{before}} &\neq KE_{\text{after}}\end{aligned}$$

8. The frame of reference is the point of view from which a system is observed.

6.7 End of chapter exercises

1. [IEB 2005/11 HG] Two friends, Ann and Lindiwe decide to race each other by swimming across a river to the other side. They swim at identical speeds relative to the water. The river has a current flowing to the east.



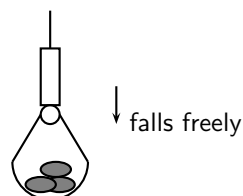
Ann heads a little west of north so that she reaches the other side directly across from the starting point. Lindiwe heads north but is carried downstream, reaching the other side downstream of Ann. Who wins the race?

- A Ann
 - B Lindiwe
 - C It is a dead heat
 - D One cannot decide without knowing the velocity of the current.
2. [SC 2001/11 HG1] A bullet fired vertically upwards reaches a maximum height and falls back to the ground.



Which **one** of the following statements is **true** with reference to the acceleration of the bullet during its motion, if air resistance is ignored? The acceleration:

- A is always downwards
 - B is first upwards and then downwards
 - C is first downwards and then upwards
 - D decreases first and then increases
3. [SC 2002/03 HG1] Thabo suspends a bag of tomatoes from a spring balance held vertically. The balance itself weighs 10 N and he notes that the balance reads 50 N. He then lets go of the balance and the balance and tomatoes fall freely. What would the reading be on the balance while falling?



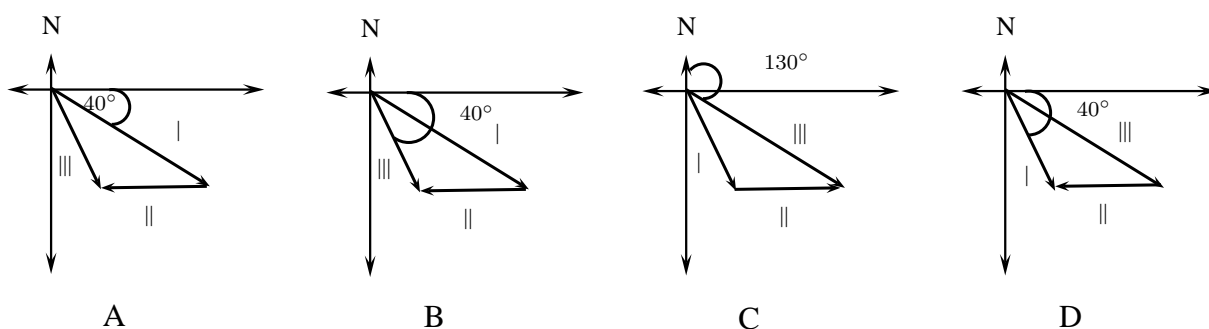
- A 50 N
 - B 40 N
 - C 10 N
 - D 0 N
4. [IEB 2002/11 HG1] Two balls, P and Q, are simultaneously thrown into the air from the same height above the ground. P is thrown vertically upwards and Q vertically downwards with the same initial speed. Which of the following is true of both balls just before they hit the ground? (Ignore any air resistance. Take downwards as the positive direction.)

	Velocity	Acceleration
A	The same	The same
B	P has a greater velocity than Q	P has a negative acceleration; Q has a positive acceleration
C	P has a greater velocity than Q	The same
D	The same	P has a negative acceleration; Q has a positive acceleration

5. [IEB 2002/11 HG1] An observer on the ground looks up to see a bird flying overhead along a straight line on bearing 130° (40° S of E). There is a steady wind blowing from east to west. In the vector diagrams below, I, II and III represent the following:

- I the velocity of the bird relative to the air
 II the velocity of the air relative to the ground
 III the resultant velocity of the bird relative to the ground

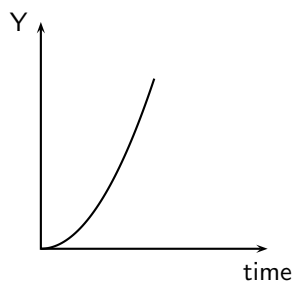
Which diagram correctly shows these three velocities?



6. [SC 2003/11] A ball X of mass m is projected vertically upwards at a speed u_x from a bridge 20 m high. A ball Y of mass $2m$ is projected vertically downwards from the same bridge at a speed of u_y . The two balls reach the water at the same speed. Air friction can be ignored.

Which of the following is true with reference to the speeds with which the balls are projected?

- A $u_x = \frac{1}{2}u_y$
 B $u_x = u_y$
 C $u_x = 2u_y$
 D $u_x = 4u_y$
7. [SC 2002/03 HG1] A stone falls freely from rest from a certain height. Which one of the following quantities could be represented on the y -axis of the graph below?



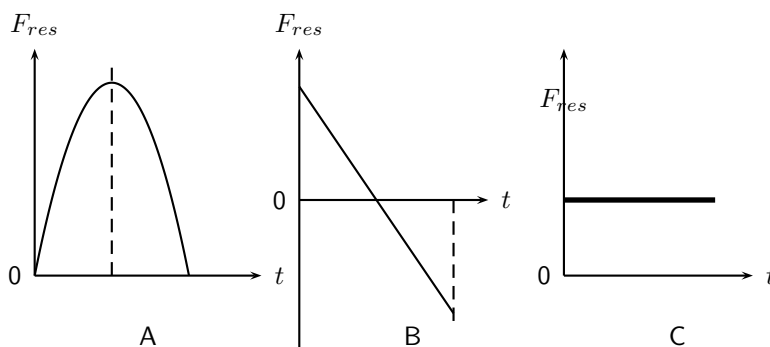
- A velocity
 B acceleration
 C momentum
 D displacement
8. A man walks towards the back of a train at $2 \text{ m}\cdot\text{s}^{-1}$ while the train moves forward at $10 \text{ m}\cdot\text{s}^{-1}$. The magnitude of the man's velocity with respect to the ground is

- A $2 \text{ m}\cdot\text{s}^{-1}$
- B $8 \text{ m}\cdot\text{s}^{-1}$
- C $10 \text{ m}\cdot\text{s}^{-1}$
- D $12 \text{ m}\cdot\text{s}^{-1}$

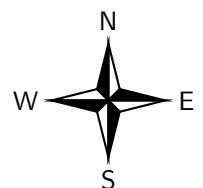
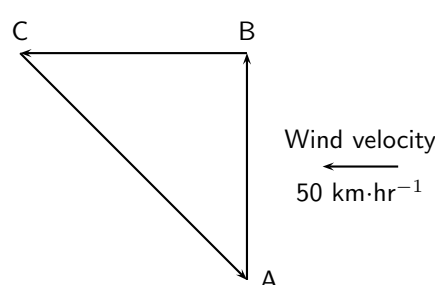
9. A stone is thrown vertically upwards and it returns to the ground. If friction is ignored, its acceleration as it reaches the highest point of its motion is
- A greater than just after it left the throwers hand.
 - B less than just before it hits the ground.
 - C the same as when it left the throwers hand.
 - D less than it will be when it strikes the ground.
10. An exploding device is thrown vertically upwards. As it reaches its highest point, it explodes and breaks up into three pieces of **equal mass**. Which one of the following combinations is possible for the motion of the three pieces if they all move in a vertical line?

	Mass 1	Mass 2	Mass 3
A	v downwards	v downwards	v upwards
B	v upwards	$2v$ downwards	v upwards
C	$2v$ upwards	v downwards	v upwards
D	v upwards	$2v$ downwards	v downwards

11. [IEB 2004/11 HG1] A stone is thrown vertically up into the air. Which of the following graphs best shows the resultant force exerted on the stone against time while it is in the air? (Air resistance is negligible.)

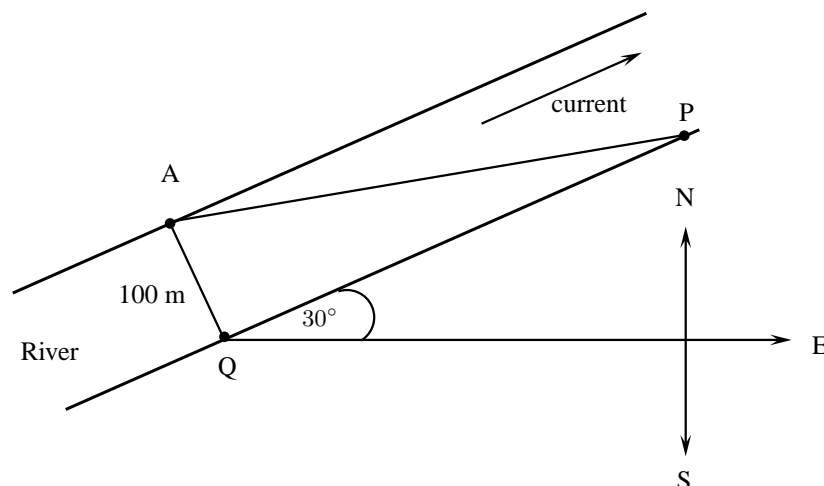


12. What is the velocity of a ball just as it hits the ground if it is thrown upward at $10 \text{ m}\cdot\text{s}^{-1}$ from a height 5 meters above the ground?
13. [IEB 2005/11 HG1] A breeze of $50 \text{ km}\cdot\text{hr}^{-1}$ blows towards the west as a pilot flies his light plane from town A to village B. The trip from A to B takes 1 h. He then turns west, flying for $\frac{1}{2}$ h until he reaches a dam at point C. He turns over the dam and returns to town A. The diagram shows his flight plan. It is not to scale.



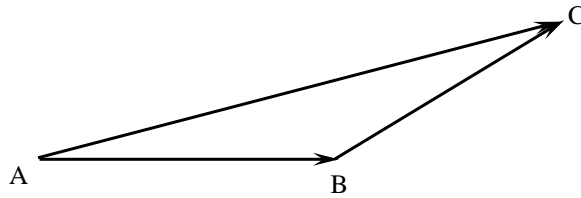
The pilot flies at the same altitude at a constant speed of $130 \text{ km}\cdot\text{h}^{-1}$ relative to the air throughout this flight.

- a Determine the magnitude of the pilot's resultant velocity from the town A to the village B.
- b How far is village B from town A?
- c What is the plane's speed relative to the ground as it travels from village B to the dam at C?
- d Determine the following, by calculation or by scale drawing:
- The distance from the village B to the dam C.
 - The displacement from the dam C back home to town A.
14. A cannon (assumed to be at ground level) is fired off a flat surface at an angle, θ above the horizontal with an initial speed of v_0 .
- What is the initial horizontal component of the velocity?
 - What is the initial vertical component of the velocity?
 - What is the horizontal component of the velocity at the highest point of the trajectory?
 - What is the vertical component of the velocity at that point?
 - What is the horizontal component of the velocity when the projectile lands?
 - What is the vertical component of the velocity when it lands?
15. [IEB 2004/11 HG1] Hailstones fall vertically on the hood of a car parked on a horizontal stretch of road. The average terminal velocity of the hailstones as they descend is $8,0 \text{ m}\cdot\text{s}^{-1}$ and each has a mass of $1,2 \text{ g}$.
- Calculate the magnitude of the momentum of a hailstone just before it strikes the hood of the car.
 - If a hailstone rebounds at $6,0 \text{ m}\cdot\text{s}^{-1}$ after hitting the car's hood, what is the magnitude of its change in momentum?
 - The hailstone is in contact with the car's hood for $0,002 \text{ s}$ during its collision with the hood of the car. What is the magnitude of the resultant force exerted on the hood if the hailstone rebounds at $6,0 \text{ m}\cdot\text{s}^{-1}$?
 - A car's hood can withstand a maximum impulse of $0,48 \text{ N}\cdot\text{s}$ without leaving a permanent dent. Calculate the minimum mass of a hailstone that will leave a dent in the hood of the car, if it falls at $8,0 \text{ m}\cdot\text{s}^{-1}$ and rebounds at $6,0 \text{ m}\cdot\text{s}^{-1}$ after a collision lasting $0,002 \text{ s}$.
16. [IEB 2003/11 HG1 - Biathlon] Andrew takes part in a biathlon race in which he first swims across a river and then cycles. The diagram below shows his points of entry and exit from the river, A and P, respectively.



During the swim, Andrew maintains a constant velocity of $1,5 \text{ m}\cdot\text{s}^{-1}$ East relative to the water. The water in the river flows at a constant velocity of $2,5 \text{ m}\cdot\text{s}^{-1}$ in a direction 30° North of East. The width of the river is 100 m.

The diagram below is a velocity-vector diagram used to determine the resultant velocity of Andrew relative to the river bed.

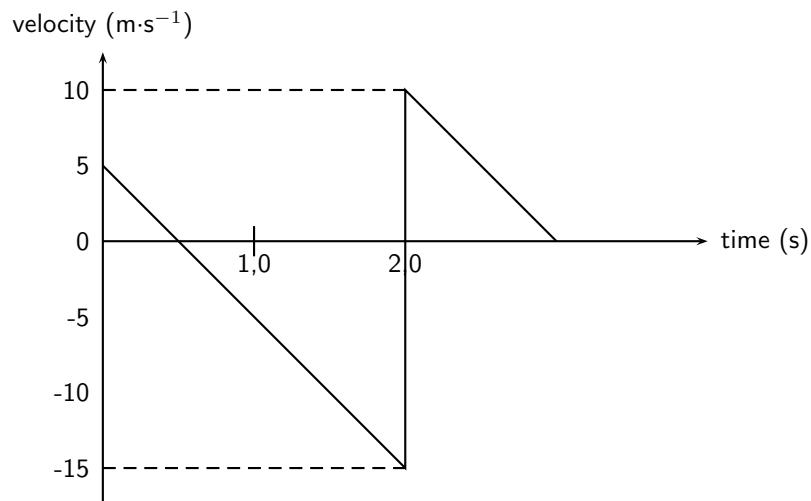


- Which of the vectors (AB, BC and AC) refer to each of the following?
 - The velocity of Andrew relative to the water.
 - The velocity of the water relative to the water bed.
 - The resultant velocity of Andrew relative to the river bed.
 - Determine the magnitude of Andrew's velocity relative to the river bed either by calculations or by scale drawing, showing your method clearly.
 - How long (in seconds) did it take Andrew to cross the river?
 - At what distance along the river bank (QP) should Peter wait with Andrew's bicycle ready for the next stage of the race?
17. [IEB 2002/11 HG1 - Bouncing Ball]

A ball bounces vertically on a hard surface after being thrown vertically up into the air by a boy standing on the ledge of a building.

Just before the ball hits the ground for the first time, it has a velocity of magnitude $15 \text{ m}\cdot\text{s}^{-1}$. Immediately, after bouncing, it has a velocity of magnitude $10 \text{ m}\cdot\text{s}^{-1}$.

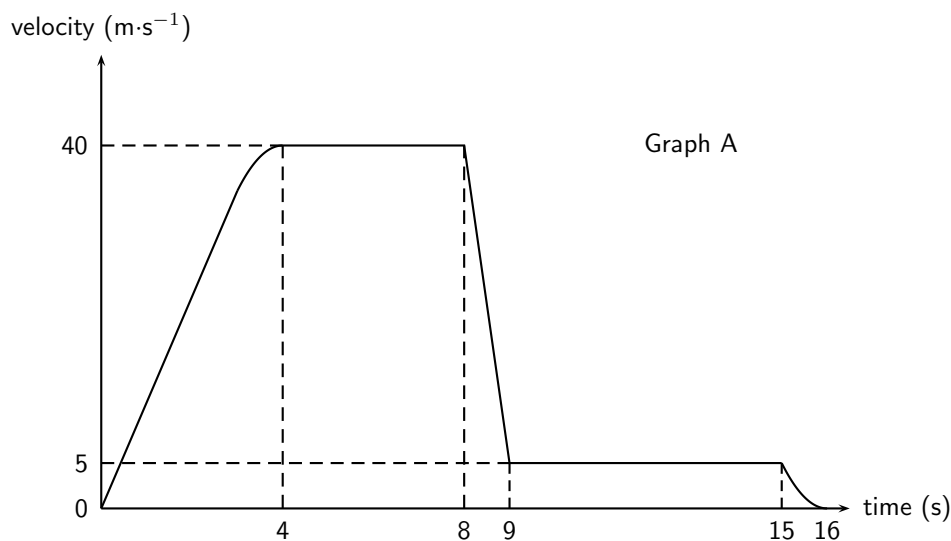
The graph below shows the velocity of the ball as a function of time from the moment it is thrown upwards into the air until it reaches its maximum height after bouncing once.



- At what velocity does the boy throw the ball into the air?
- What can be determined by calculating the gradient of the graph during the first two seconds?
- Determine the gradient of the graph over the first two seconds. State its units.
- How far below the boy's hand does the ball hit the ground?
- Use an equation of motion to calculate how long it takes, from the time the ball was thrown, for the ball to reach its maximum height after bouncing.
- What is the position of the ball, measured from the boy's hand, when it reaches its maximum height after bouncing?

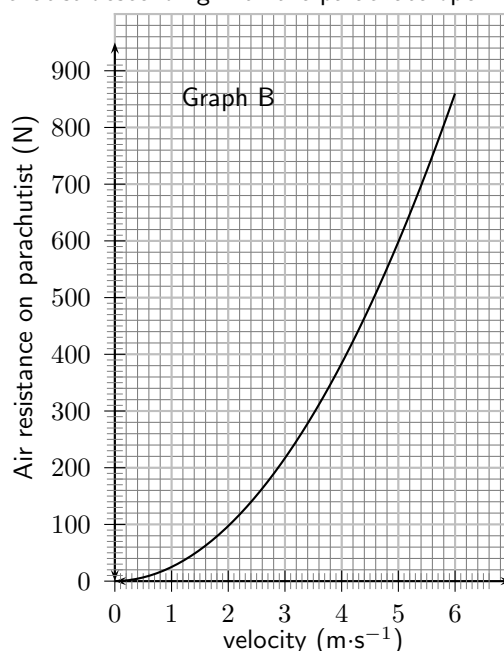
18. [IEB 2001/11 HG1] - **Free Falling?**

A parachutist steps out of an aircraft, flying high above the ground. She falls for the first 8 seconds before opening her parachute. A graph of her velocity is shown in Graph A below.



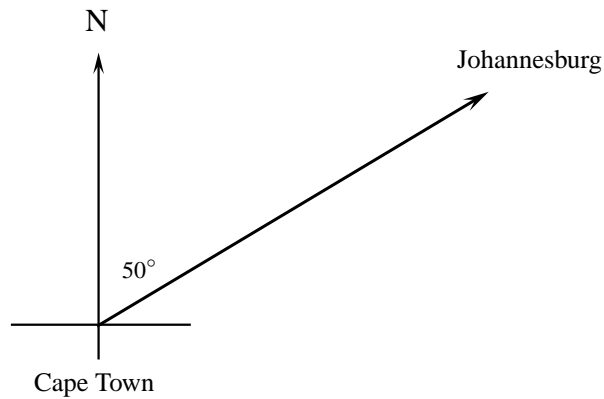
- a Use the information from the graph to calculate an approximate height of the aircraft when she stepped out of it (to the nearest 10 m).
- b What is the magnitude of her velocity during her descent with the parachute fully open?

The air resistance acting on the parachute is related to the speed at which the parachutist descends. Graph B shows the relationship between air resistance and velocity of the parachutist descending with the parachute open.

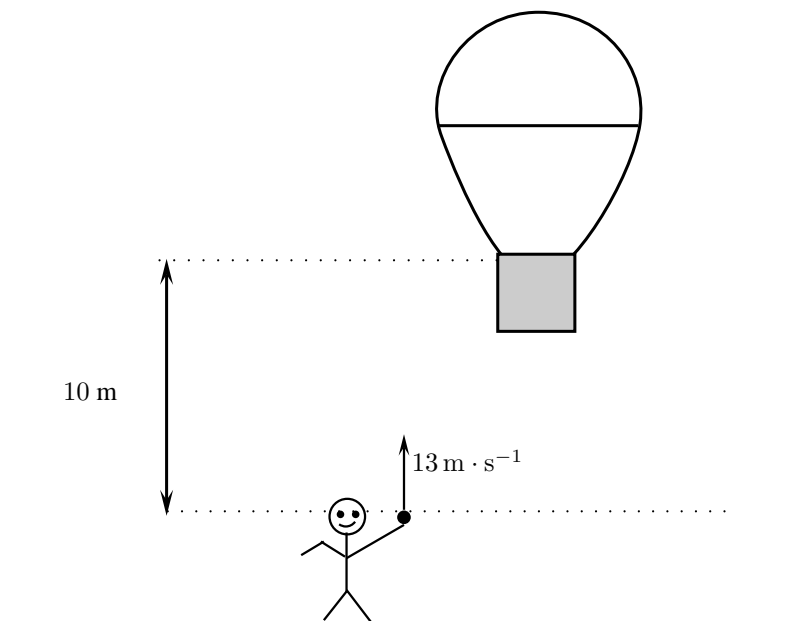


- c Use Graph B to find the magnitude of the air resistance on her parachute when she was descending with the parachute open.
- d Assume that the mass of the parachute is negligible. Calculate the mass of the parachutist showing your reasoning clearly.

19. An aeroplane travels from Cape Town and the pilot must reach Johannesburg, which is situated 1300 km from Cape Town on a bearing of 50° in 5 hours. At the height at which the plane flies, a wind is blowing at $100 \text{ km}\cdot\text{hr}^{-1}$ on a bearing of 130° for the whole trip.



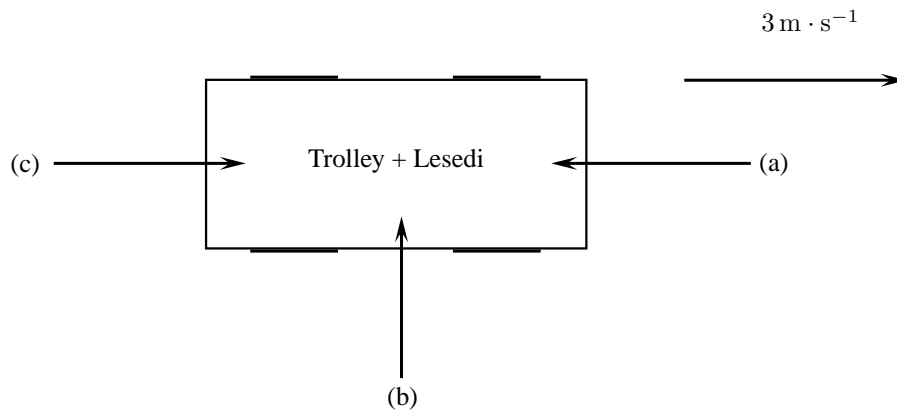
- a Calculate the magnitude of the average resultant velocity of the aeroplane, in $\text{km}\cdot\text{hr}^{-1}$, if it is to reach its destination on time.
- b Calculate the average velocity, in $\text{km}\cdot\text{hr}^{-1}$, in which the aeroplane should be travelling in order to reach Johannesburg in the prescribed 5 hours. Include a labelled, rough vector diagram in your answer.
(If an accurate scale drawing is used, a scale of $25 \text{ km}\cdot\text{hr}^{-1} = 1 \text{ cm}$ must be used.)
20. Niko, in the basket of a hot-air balloon, is stationary at a height of 10 m above the level from where his friend, Bongi, will throw a ball. Bongi intends throwing the ball upwards and Niko, in the basket, needs to **descend** (move downwards) to catch the ball at its maximum height.



Bongi throws the ball upwards with a velocity of $13 \text{ m}\cdot\text{s}^{-1}$. Niko starts his descent at the same instant the ball is thrown upwards, by letting air escape from the balloon, causing it to accelerate downwards. Ignore the effect of air friction on the ball.

- a Calculate the maximum height reached by the ball.
- b Calculate the magnitude of the minimum average acceleration the balloon must have in order for Niko to catch the ball, if it takes 1,3 s for the ball to reach its maximum height.
21. Lesedi (mass 50 kg) sits on a massless trolley. The trolley is travelling at a constant speed of $3 \text{ m}\cdot\text{s}^{-1}$. His friend Zola (mass 60 kg) jumps on the trolley with a velocity of $2 \text{ m}\cdot\text{s}^{-1}$. What is the final velocity of the combination (Lesedi, Zola and trolley) if Zola jumps on the trolley from
- a the front
- b behind

c the side
(Ignore all kinds of friction)



Chapter 7

Mechanical Properties of Matter - Grade 12

7.1 Introduction

In this chapter we will look at some mechanical (physical) properties of various materials that we use. The mechanical properties of a material are those properties that are affected by forces being applied to the material. These properties are important to consider when we are constructing buildings, structures or modes of transport like an aeroplane.

7.2 Deformation of materials

7.2.1 Hooke's Law

Deformation (change of shape) of a solid is caused by a force that can either be compressive or tensile when applied in one direction (plane). Compressive forces try to compress the object (make it smaller or more compact) while tensile forces try to tear it apart. We can study these effects by looking at what happens when you compress or expand a spring.

Hooke's Law relates the restoring force of a spring to its displacement from equilibrium length.

The equilibrium length of a spring is its length when no forces are applied to it. When a force is applied to a spring, e.g., by attaching a weight to one end, the spring will expand and become longer. The difference between the new length and the equilibrium length is the displacement.

Historical Note: Hooke's Law

Hooke's law is named after the seventeenth century physicist Robert Hooke who discovered it in 1660 (18 July 1635 - 3 March 1703).

**Definition: Hooke's Law**

In an elastic spring, the extension varies linearly with the force applied.

$$F = -kx$$

where F is the restoring force in newtons (N), k is the spring constant in $N \cdot m^{-1}$ and x is the displacement of the spring from its equilibrium length in metres (m).

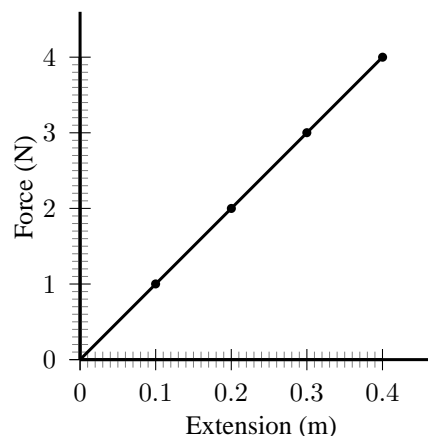


Figure 7.1: Hooke's Law - the relationship between a spring's restoring force and its displacement from equilibrium length.

Activity :: Experiment : Hooke's Law

Aim:

Verify Hooke's Law.

Apparatus:

- weights
- spring
- ruler

Method:

1. Set up a spring vertically in such a way that you are able to hang weights from it.
2. Measure the equilibrium length, x_0 , of the spring (i.e. the length of the spring when nothing is attached to it).
3. Measure the extension of the spring for a range of different weights. Note: the extension is the difference between the spring's equilibrium length and the new length when a weight is attached to it, $x - x_0$.
4. Draw a table of force (weight) in newtons and corresponding extension.
5. Draw a graph of force versus extension for your experiment.

Conclusions:

1. What do you observe about the relationship between the applied force and the extension?
 2. Determine the gradient (slope) of the graph.
 3. Now calculate the spring constant for your spring.
-



Worked Example 43: Hooke's Law I

Question: A spring is extended by 7 cm by a force of 56 N. Calculate the spring constant for this spring.

Answer

$$F = -kx$$

$$56 = -k \times 0,07$$

$$k = \frac{-56}{0,07}$$

$$= -800 \text{ N} \cdot \text{m}^{-1}$$



Worked Example 44: Hooke's Law II

Question: A spring of length 20cm stretches to 24cm when a load of 0,6N is applied to it.

1. Calculate the spring constant for the spring.
2. Determine the extension of the spring if a load of 0,5N is applied to it.

Answer

Step 1 : Determine what information you have:

We know:

$$F = 0,6 \text{ N}$$

The equilibrium spring length is 20 cm

The expanded spring length is 24 cm

Step 2 : Use Hooke's Law to find the spring constant

First we need to calculate the displacement of the spring from its equilibrium length:

$$x = 24 \text{ cm} - 20 \text{ cm}$$

$$= 4 \text{ cm}$$

$$= 0,04 \text{ m}$$

Now use Hooke's Law to find the spring constant:

$$F = -kx$$

$$0,6 = -k \cdot 0,04$$

$$k = -15 \text{ N} \cdot \text{m}^{-1}$$

Step 3 : The second part of the question asks us to find the spring's extension if a 0,5 N load is attached to it. We have:

$$F = 0,5 \text{ N}$$

We know from the first part of the question that

$$k = -15 \text{ N} \cdot \text{m}^{-1}$$

So, using Hooke's Law:

$$\begin{aligned}
 F &= -kx \\
 x &= -\frac{F}{k} \\
 &= -\frac{0,5}{-15} \\
 &= 0,033 \text{ m} \\
 &= 3,3 \text{ cm}
 \end{aligned}$$



Worked Example 45: Hooke's Law III

Question: A spring has a spring constant of $-400 \text{ N}\cdot\text{m}^{-1}$. By how much will it stretch if a load of 50 N is applied to it?

Answer

$$\begin{aligned}
 F &= -kx \\
 50 &= -(-400)x
 \end{aligned}$$

$$\begin{aligned}
 x &= \frac{50}{400} \\
 &= 0,125 \text{ m} \\
 &= 12,5 \text{ cm}
 \end{aligned}$$

7.2.2 Deviation from Hooke's Law

We know that if you have a small spring and you pull it apart too much it stops 'working'. It bends out of shape and loses its springiness. When this happens, Hooke's Law no longer applies, the spring's behaviour deviates from Hooke's Law.

Depending on what type of material we are dealing with, the manner in which it deviates from Hooke's Law is different. We give classify materials by this deviation. The following graphs show the relationship between force and extension for different materials and they all deviate from Hooke's Law. Remember that a straight line show proportionality so as soon as the graph is no longer a straight line, Hooke's Law no longer applies.

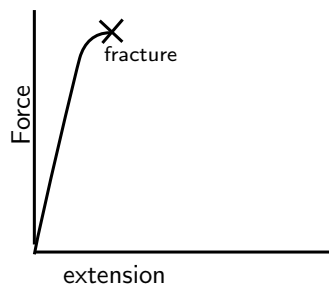
Brittle material

Figure 7.2: A hard, brittle substance

This graph shows the relationship between force and extension for a brittle, but strong material. Note that there is very little extension for a large force but then the material suddenly fractures. Brittleness is the property of a material that makes it break easily without bending.

Have you ever dropped something made of glass and seen it shatter? Glass does this because it is brittle.

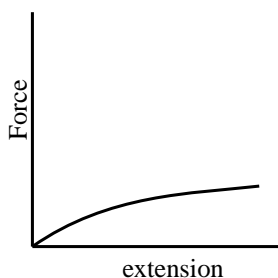
Plastic material

Figure 7.3: A plastic material's response to an applied force.

Here the graph shows the relationship between force and extension for a plastic material. The material extends under a small force but it does not fracture easily, and it does not return to its original length when the force is removed.

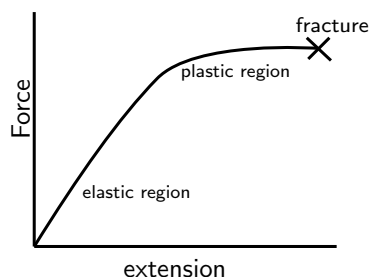
Ductile material

Figure 7.4: A ductile substance.

In this graph the relationship between force and extension is for a material that is ductile. The material shows plastic behaviour over a range of forces before the material finally fractures.

Ductility is the ability of a material to be stretched into a new shape without breaking. Ductility is one of the characteristic properties of metals.

A good example of this is aluminium, many things are made of aluminium. Aluminium is used for making everything from cooldrink cans to aeroplane parts and even engine blocks for cars. Think about squashing and bending a cooldrink can.

Brittleness is the opposite of ductility.

When a material reaches a point where Hooke's Law is no longer valid, we say it has reached its *limit of proportionality*. After this point, the material will not return to its original shape after the force has been removed. We say it has reached its *elastic limit*.



Definition: Elastic limit

The elastic limit is the point beyond which permanent deformation takes place.



Definition: Limit of proportionality

The limit of proportionality is the point beyond which Hooke's Law is no longer obeyed.



Exercise: Hooke's Law and deformation of materials

1. What causes deformation?
2. Describe Hooke's Law in words and mathematically.
3. List similarities and differences between ductile, brittle and plastic (polymeric) materials, with specific reference to their force-extension graphs.
4. Describe what is meant by the *elastic limit*.
5. Describe what is meant by the *limit of proportionality*.
6. A spring of length 15 cm stretches to 27 cm when a load of 0,4 N is applied to it.
 - (a) Calculate the spring constant for the spring.
 - (b) Determine the extension of the spring if a load of 0,35 N is applied to it.
7. A spring has a spring constant of $-200 \text{ N}\cdot\text{m}^{-1}$. By how much will it stretch if a load of 25 N is applied to it?
8. A spring of length 20 cm stretches to 24 cm when a load of 0,6 N is applied to it.
 - (a) Calculate the spring constant for the spring.
 - (b) Determine the extension of the spring if a load of 0,8 N is applied to it.

7.3 Elasticity, plasticity, fracture, creep

7.3.1 Elasticity and plasticity

Materials are classified as plastic or elastic depending on how they respond to an applied force. It is important to note that plastic substances are not necessarily a type of plastic (polymer) they only behave like plastic. Think of them as being like plastic which you will be familiar with.

A rubber band is a material that has elasticity. It returns to its original shape after an applied force is removed, providing that the material is not stretched beyond its elastic limit.

Plasticine is an example of a material that is plastic. If you flatten a ball of plasticine, it will stay flat. A plastic material does not return to its original shape after an applied force is removed.

- Elastic materials return to their original shape.
- Plastic materials deform easily and do not return to their original shape.

7.3.2 Fracture, creep and fatigue

Some materials are neither plastic nor elastic. These substances will break or fracture when a large enough force is applied to them. The brittle glass we mentioned earlier is an example.

Creep occurs when a material deforms over a long period of time because of an applied force. An example of creep is the bending of a shelf over time when a heavy object is put on it. Creep may eventually lead to the material fracturing. The application of heat may lead to an increase in creep in a material.

Fatigue is similar to creep. The difference between the two is that fatigue results from the force being applied and then removed repeatedly over a period of time. With metals this results in failure because of metal fatigue.

- Fracture is an abrupt breaking of the material.
- Creep is a slow deformation process due to a continuous force over a long time.
- Fatigue is weakening of the material due to short forces acting many many times.



Exercise: Elasticity, plasticity, fracture and creep

1. List the similarities and differences between elastic and plastic deformation.
 2. List the similarities and differences between creep and fracture as modes of failure in material.
-

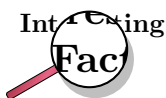
7.4 Failure and strength of materials

7.4.1 The properties of matter

The strength of a material is defined as the stress (the force per unit cross-sectional area) that it can withstand. Strength is measured in newtons per square metre ($N \cdot m^{-2}$).

Stiffness is a measure of how flexible a material is. In Science we measure the stiffness of a material by calculating its Young's Modulus. The Young's modulus is a ratio of how much it bends to the load applied to it. Stiffness is measure in newtons per metre ($N \cdot m^{-1}$).

Hardness of a material can be measured by determining what force will cause a permanent deformation in the material. Hardness can also be measured using a scale like Mohs hardness scale. On this scale, diamond is the hardest at 10 and talc is the softest at 1.



Remembering that the Mohs scale is the hardness scale and that the softest substance is talc will often come in handy for general knowledge quizzes.

The *toughness* of a material is a measure of how it can resist breaking when it is stressed. It is scientifically defined as the amount of energy that a material can absorb before fracturing.

A ductile material is a substance that can undergo large plastic deformation without fracturing. Many metals are very ductile and they can be drawn into wires, e.g. copper, silver, aluminium and gold.

A *malleable* material is a substance that can easily undergo plastic deformation by hammering or rolling. Again, metals are malleable substances, e.g. copper can be hammered into sheets and aluminium can be rolled into aluminium foil.

A brittle material fractures with very little or no plastic deformation. Glassware and ceramics are brittle.

7.4.2 Structure and failure of materials

Many substances fail because they have a weakness in their atomic structure. There are a number of problems that can cause these weaknesses in structure. These are vacancies, dislocations, grain boundaries and impurities.

Vacancies occur when there are spaces in the structure of a crystalline solid. These vacancies cause weakness and such materials often fail at these places. Think about bricks in a wall, if you started removing bricks the wall would get weaker.

Dislocations result in weakened bonding between layers of atoms in a crystal lattice and this creates a critical boundary. If sufficient force is applied along the boundary, it can break the weakened bonds, allowing the two sides of the crystal to slide against one another. The two pieces of the crystal keep their shape and structure.

Impurities in a crystal structure can cause a weak region in the crystal lattice around the impurity. Like vacancies, the substance often fail from these places in the lattice. This you can think of as bricks in a wall which don't fit properly, they are the wrong kind of bricks (atoms) to make the structure strong.

7.4.3 Controlling the properties of materials

There are a number of processes that can be used to make materials less likely to fail. We shall look at a few methods in this section.

Cold working

Cold working is a process in which a metal is *strengthened* by repeatedly being reshaped. This is carried out at a temperature below the melting point of the metal. The repeated shaping of the metal results in dislocations which then prevent restrict the motion of dislocations in the metal. Cold working increases the strength of the metal but in so doing, the metal loses its ductility. We say the metal is *work-hardened*.

Annealing

Annealing is a process of heating and cooling a material to relax the crystal structure and reduce weakness due to impurities and structural flaws. During annealing, the material is heated to a high temperature that is below the material's melting point. At a sufficiently high temperature, atoms with weakened bonds can rearrange themselves into a stronger structure. Slowly cooling the material ensures that the atoms will remain in these stronger locations. Annealing is often used before cold working.

Introduction of Impurities

Most pure metals are relatively weak because dislocations can move easily within them. However, if impurities are added to a metal (e.g., carbon is added to an iron sample), they can disturb the regular structure of the metal and so prevent dislocations from spreading. This makes the metal stronger.

Alloying

An alloy is a mixture of a metal with other substances. In other words, alloying involves adding impurities to a metal sample. The other substances can be metal or non-metal. An alloy often has properties that are very different to the properties of the substances from which it is made. The added substances strengthen the metal by preventing dislocations from spreading. Ordinary steel is an alloy of iron and carbon. The carbon impurities trap dislocations. There are many types of steel that also include other metals with iron and carbon. Brass is an alloy of copper and Zinc. Bronze is an alloy of copper and tin. Gold and silver that is used in coins or jewellery are also alloyed.

Tempering

Tempering is a process in which a metal is melted then quickly cooled. The rapid cooling is called quenching. Usually tempering is done a number of times before a metal has the correct properties that are needed for a particular application.

Sintering

Sintering is used for making ceramic objects among other things. In this process the substance is heated so that its particles stick together. It is used with substances that have a very high melting point. The resulting product is often very pure and it is formed in the process into the shape that is wanted. Unfortunately, sintered products are brittle.

7.4.4 Steps of Roman Swordsmithing

- Purifying the iron ore.
- Heating the iron blocks in a furnace with charcoal.
- Hammering and getting into the needed shape. The smith used a hammer to pound the metal into blade shape. He usually used tongs to hold the iron block in place.
- Reheating. When the blade cooled, the smith reheated it to keep it workable. While reheated and hammered repeatedly.
- *Quenching* which involved the process of white heating and cooling in water. Quenching made the blade harder and stronger. At the same time it made the blade quite brittle, which was a considerable problem for the sword smiths.
- *Tempering* was then done to avoid brittleness the blade was tempered. In another words it was reheated a final time to a very specific temperature. How the Romans do balanced the temperature? The smith was guided only by the blade's color and his own experience.



Exercise: Failure and strength of materials

1. List the similarities and differences between the brittle and ductile modes of failure.

2. What is meant by the following terms:
 - (a) vacancies
 - (b) dislocations
 - (c) impurities
 - (d) grain boundaries
 3. What four terms can be used to describe a material's mechanical properties?
 4. What is meant by the following:
 - (a) cold working
 - (b) annealing
 - (c) tempering
 - (d) introduction of impurities
 - (e) alloying
 - (f) sintering
-

7.5 Summary

1. Hooke's Law gives the relationship between the extension of a spring and the force applied to it. The law says they are proportional.
2. Materials can be classified as plastic or elastic depending on how they respond to an applied force.
3. Materials can fracture or undergo creep or fatigue when forces are applied to them.
4. Materials have the following mechanical properties to a greater or lesser degree: strength, hardness, ductility, malleability, brittleness, stiffness.
5. Materials can be weakened by have the following problems in their crystal lattice: vacancies, dislocations, impurities, difference in grain size.
6. Materials can have their mechanical properties improved by one or more of the following processes: cold working, annealing, adding impurities, tempering, sintering.

7.6 End of chapter exercise

1. State Hooke's Law in words.
2. What do we mean by the following terms with respect to Hooke's Law?
 - (a) elastic limit
 - (b) limit of proportionality
3. A spring is extended by 18 cm by a force of 90 N. Calculate the spring constant for this spring.
4. A spring of length 8 cm stretches to 14 cm when a load of 0,8 N is applied to it.
 - (a) Calculate the spring constant for the spring.
 - (b) Determine the extension of the spring if a load of 0,7 N is applied to it.
5. A spring has a spring constant of -150 N.m^{-1} . By how much will it stretch if a load of 80 N is applied to it?
6. What do we mean by the following terms when speaking about properties of materials?
 - (a) hardness

- (b) toughness
 - (c) ductility
 - (d) malleability
 - (e) stiffness
 - (f) strength
7. What is Young's modulus?
 8. In what different ways can we improve the material properties of substances?
 9. What is a metal alloy?
 10. What do we call an alloy of:
 - (a) iron and carbon
 - (b) copper and zinc
 - (c) copper and tin
 11. Do some research on what added substances can do to the properties of steel. Present your findings in a suitable table.

Chapter 8

Work, Energy and Power - Grade 12

8.1 Introduction

Imagine a vendor carrying a basket of vegetables on her head. Is she doing any work? One would definitely say yes! However, in Physics she is not doing any work! Again, imagine a boy pushing against a wall? Is he doing any work? We can see that his muscles are contracting and expanding. He may even be sweating. But in Physics, he is not doing any work!

If the vendor is carrying a very heavy load for a long distance, we would say she has a lot of energy. By this, we mean that she has a lot of stamina. If a car can travel very fast, we describe the car as powerful. So, there is a link between power and speed. However, power means something different in Physics. This chapter describes the links between work, energy and power and what these mean in Physics.

You will learn that work and energy are closely related. You shall see that the energy of an object is its capacity to do work and doing work is the process of transferring energy from one object or form to another. In other words,

- an object with lots of energy can do lots of work.
- when work is done, energy is lost by the object doing work and gained by the object on which the work is done.

Lifting objects or throwing them requires that you do work on them. Even making electricity flow requires that something do work. Something must have energy and transfer it through doing work to make things happen.

8.2 Work

**Definition: Work**

When a force exerted on an object causes it to move, work is done on the object (except if the force and displacement are at right angles to each other).

This means that in order for work to be done, an object must be moved a distance d by a force F , such that there is some non-zero component of the force in the direction of the displacement. Work is calculated as:

$$W = F \cdot \Delta x \cos \theta. \quad (8.1)$$

where F is the applied force, Δx is the displacement of the object and θ is the angle between the applied force and the direction of motion.

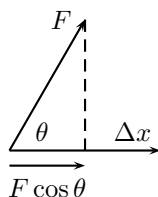


Figure 8.1: The force F causes the object to be displaced by Δx at angle θ .

It is very important to note that for work to be done there must be a component of the applied force in the direction of motion. Forces perpendicular to the direction of motion do no work.

For example work is done on the object in Figure 8.2,

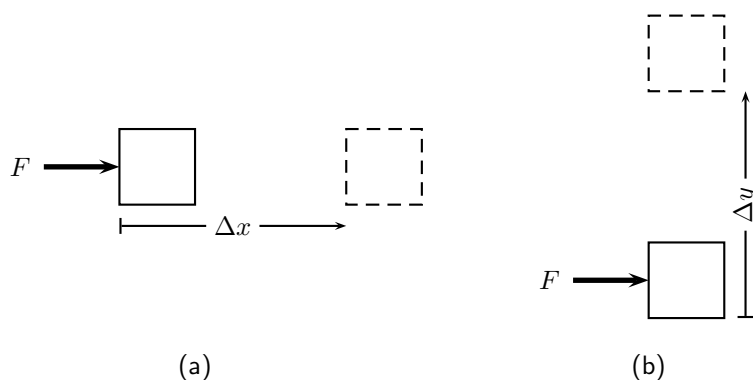


Figure 8.2: (a) The force F causes the object to be displaced by Δx in the same direction as the force. $\theta = 0^\circ$ and $\cos \theta = 1$. Work is done in this situation. (b) A force F is applied to the object. The object is displaced by Δy at right angles to the force. $\theta = 90^\circ$ and $\cos \theta = 0$. Work is not done in this situation.

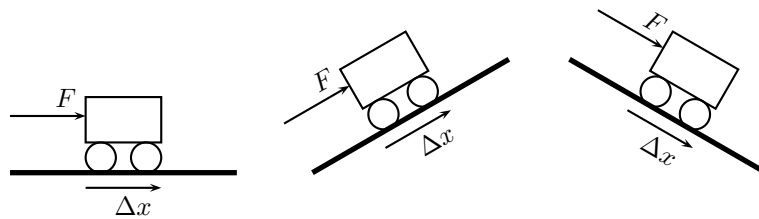
Activity :: Investigation : Is work done?

Decide whether or not work is done in the following situations. Remember that for work to be done a force must be applied in the direction of motion and there must be a displacement. Give reasons for your answer.

1. Max pushes against a wall and becomes tired.
2. A book falls off a table and free falls to the ground.
3. A rocket accelerates through space.
4. A waiter holds a tray full of meals above his head with one arm and carries it straight across the room at constant speed. (Careful! This is a tricky question.)



Important: The Meaning of θ : The angle θ is the angle between the force vector and the displacement vector. In the following situations, $\theta = 0^\circ$.



As with all physical quantities, work must have units. Following from the definition, work is measured in N·m. The name given to this combination of S.I. units is the joule (symbol J).



Definition: Joule

1 joule is the work done when an object is moved 1 m under the application of a force of 1 N in the direction of motion.

The work done by an object can be positive or negative. Since force (F_{\parallel}) and displacement (s) are both vectors, the result of the above equation depends on their directions:

- If F_{\parallel} acts in the same direction as the motion then positive work is being done. In this case the object on which the force is applied gains energy.
- If the direction of motion and F_{\parallel} are opposite, then negative work is being done. This means that energy is transferred in the opposite direction. For example, if you try to push a car uphill by applying a force up the slope and instead the car rolls down the hill you are doing negative work on the car. Alternatively, the car is doing positive work on you!



Important: The everyday use of the word "work" differs from the physics use. In physics, only the component of the applied force that is parallel to the motion does work on an object. So, for example, a person holding up a heavy book does no work on the book.



Worked Example 46: Calculating Work Done I

Question: If you push a box 20 m forward by applying a force of 15 N in the forward direction, what is the work you have done on the box?

Answer

Step 1 : Analyse the question to determine what information is provided

- The force applied is $F=15$ N.
- The distance moved is $s=20$ m.
- The applied force and distance moved are in the same direction. Therefore, $F_{\parallel}=15$ N.

These quantities are all in the correct units, so no unit conversions are required.

Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the box. We know from the definition that work done is $W = F_{\parallel}s$

Step 3 : Next we substitute the values and calculate the work done

$$\begin{aligned} W &= F_{\parallel}s \\ &= (15 \text{ N})(20 \text{ m}) \\ &= 300 \text{ J} \end{aligned}$$

Remember that the answer must be *positive* as the applied force and the motion are in the same direction (forwards). In this case, you (the pusher) lose energy, while the box gains energy.



Worked Example 47: Calculating Work Done II

Question: What is the work done by you on a car, if you try to push the car up a hill by applying a force of 40 N directed up the slope, but it slides downhill 30 cm?

Answer

Step 1 : Analyse the question to determine what information is provided

- The force applied is $F=40 \text{ N}$
- The distance moved is $s=30 \text{ cm}$. This is expressed in the wrong units so we must convert to the proper S.I. units (meters):

$$\begin{aligned} 100 \text{ cm} &= 1 \text{ m} \\ 1 \text{ cm} &= \frac{1}{100} \text{ m} \\ \therefore 30 \times 1 \text{ cm} &= 30 \times \frac{1}{100} \text{ m} \\ &= \frac{30}{100} \text{ m} \\ &= 0,3 \text{ m} \end{aligned}$$

- The applied force and distance moved are in opposite directions. Therefore, if we take $s=0.3 \text{ m}$, then $F_{\parallel}=-40 \text{ N}$.

Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the car by you. We know that work done is $W = F_{\parallel}s$

Step 3 : Substitute the values and calculate the work done

Again we have the applied force and the distance moved so we can proceed with calculating the work done:

$$\begin{aligned} W &= F_{\parallel}s \\ &= (-40 \text{ N})(0.3 \text{ m}) \\ &= -12 \text{ J} \end{aligned}$$

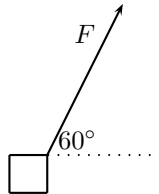
Note that the answer must be *negative* as the applied force and the motion are in opposite directions. In this case the car does work on the person trying to push.

What happens when the applied force and the motion are not parallel? If there is an angle between the direction of motion and the applied force then to determine the work done we have to calculate the *component* of the applied force *parallel* to the direction of motion. Note that this means a force perpendicular to the direction of motion can do no work.



Worked Example 48: Calculating Work Done III

Question: Calculate the work done on a box, if it is pulled 5 m along the ground by applying a force of $F=10\text{ N}$ at an angle of 60° to the horizontal.



Answer

Step 1 : Analyse the question to determine what information is provided

- The force applied is $F=10\text{ N}$
- The distance moved is $s=5\text{ m}$ along the ground
- The angle between the applied force and the motion is 60°

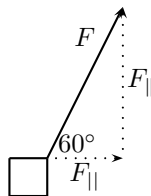
These quantities are in the correct units so we do not need to perform any unit conversions.

Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the box.

Step 3 : Calculate the component of the applied force in the direction of motion

Since the force and the motion are not in the same direction, we must first calculate the component of the force in the direction of the motion.



From the force diagram we see that the component of the applied force parallel to the ground is

$$\begin{aligned} F_{\parallel} &= F \cdot \cos(60^\circ) \\ &= 10\text{ N} \cdot \cos(60^\circ) \\ &= 5\text{ N} \end{aligned}$$

Step 4 : Substitute and calculate the work done

Now we can calculate the work done on the box:

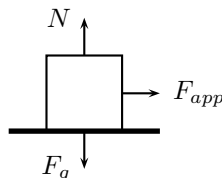
$$\begin{aligned} W &= F_{\parallel}s \\ &= (5\text{ N})(5\text{ m}) \\ &= 25\text{ J} \end{aligned}$$

Note that the answer is positive as the component of the force F_{\parallel} is in the same direction as the motion.

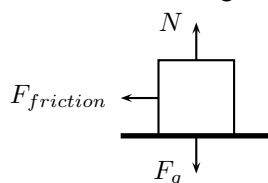


Exercise: Work

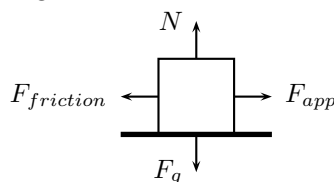
1. A 10 N force is applied to push a block across a friction free surface for a displacement of 5.0 m to the right. The block has a weight of 20 N. Determine the work done by the following forces: normal force, weight, applied force.



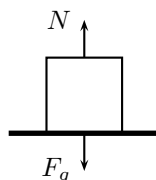
2. A 10 N frictional force slows a moving block to a stop after a displacement of 5.0 m to the right. The block has a weight of 20 N. Determine the work done by the following forces: normal force, weight, frictional force.



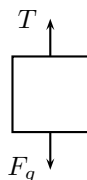
3. A 10 N force is applied to push a block across a frictional surface at constant speed for a displacement of 5.0 m to the right. The block has a weight of 20 N and the frictional force is 10 N. Determine the work done by the following forces: normal force, weight, frictional force.



4. A 20 N object is sliding at constant speed across a friction free surface for a displacement of 5 m to the right. Determine if there is any work done.



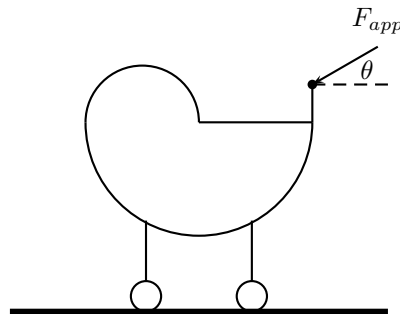
5. A 20 N object is pulled upward at constant speed by a 20 N force for a vertical displacement of 5 m. Determine if there is any work done.



6. Before beginning its descent, a roller coaster is always pulled up the first hill to a high initial height. Work is done on the roller coaster to achieve this initial height. A coaster designer is considering three different incline angles **of the hill** at which to drag the 2 000 kg car train to the top of the 60 m high hill. In each case, the force applied to the car will be applied parallel to the hill. Her critical question is: which angle would require the least work? Analyze the data, determine the work done in each case, and answer this critical question.

Angle of Incline	Applied Force	Distance	Work
35°	1.1×10^4 N	100 m	
45°	1.3×10^4 N	90 m	
55°	1.5×10^4 N	80 m	

7. Big Bertha carries a 150 N suitcase up four flights of stairs (a total height of 12 m) and then pushes it with a horizontal force of 60 N at a constant speed of $0.25 \text{ m}\cdot\text{s}^{-1}$ for a horizontal distance of 50 m on a frictionless surface. How much work does Big Bertha do on the suitcase during this entire trip?
8. A mother pushes down on a pram with a force of 50 N at an angle of 30° . The pram is moving on a frictionless surface. If the mother pushes the pram for a horizontal distance of 30 m, how much does she do on the pram?



9. How much work is done by an applied force to raise a 2 000 N lift 5 floors vertically at a constant speed? Each floor is 5 m high.
10. A student with a mass of 60 kg runs up three flights of stairs in 15 s, covering a vertical distance of 10 m. Determine the amount of work done by the student to elevate her body to this height. Assume that her speed is constant.

8.3 Energy

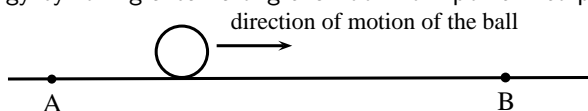
8.3.1 External and Internal Forces

In Grade 10, you saw that mechanical energy was conserved in the absence of external forces. It is important to know whether a force is an internal force or an external force in the system, because this is related to whether the force can change an object's total mechanical energy when it does work on an object.

When an external force (for example friction, air resistance, applied force) does work on an object, the total mechanical energy ($KE + PE$) of that object changes. If positive work is done, then the object will gain energy. If negative work is done, then the object will lose energy. The gain or loss in energy can be in the form of potential energy, kinetic energy, or both. However, the work which is done is equal to the change in mechanical energy of the object.

Activity :: Investigations : External Forces

We can investigate the effect of external forces on an object's total mechanical energy by rolling a ball along the floor from point A to point B.



Find a nice smooth surface (e.g. a highly polished floor), mark off two positions, A and B, and roll the ball between them.

The total mechanical energy of the ball, at each point, is the sum of its kinetic energy (KE) and gravitational potential energy (PE):

$$\begin{aligned}
 E_{\text{total,A}} &= KE_A + PE_A & E_{\text{total,B}} &= KE_B + PE_B \\
 &= \frac{1}{2}mv_A^2 + mgh_A & &= \frac{1}{2}mv_B^2 + mgh_B \\
 &= \frac{1}{2}mv_A^2 + mg(0) & &= \frac{1}{2}mv_B^2 + mg(0) \\
 &= \frac{1}{2}mv_A^2 & &= \frac{1}{2}mv_B^2
 \end{aligned}$$

In the absence of friction and other external forces, the ball should slide along the floor and its speed should be *the same* at positions A and B. Since there are no external forces acting on the ball, its total mechanical energy at points A and B are equal.

$$\begin{aligned}
 v_A &= v_B \\
 \frac{1}{2}mv_A^2 &= \frac{1}{2}mv_B^2 \\
 E_{\text{total,A}} &= E_{\text{total,B}}
 \end{aligned}$$

Now, let's investigate what happens when there is friction (an *external force*) acting on the ball.

Roll the ball along a rough surface or a carpeted floor. What happens to the speed of the ball at point A compared to point B?

If the surface you are rolling the ball along is very rough and provides a large external frictional force, then the ball should be moving much slower at point B than at point A.

Let's now compare the total mechanical energy of the ball at points A and B:

$$\begin{aligned}
 E_{\text{total,A}} &= KE_A + PE_A & E_{\text{total,B}} &= KE_B + PE_B \\
 &= \frac{1}{2}mv_A^2 + mgh_A & &= \frac{1}{2}mv_B^2 + mgh_B \\
 &= \frac{1}{2}mv_A^2 + mg(0) & &= \frac{1}{2}mv_B^2 + mg(0) \\
 &= \frac{1}{2}mv_A^2 & &= \frac{1}{2}mv_B^2
 \end{aligned}$$

However, in this case, $v_A \neq v_B$ and therefore $E_{\text{total,A}} \neq E_{\text{total,B}}$. Since

$$\begin{aligned}
 v_A &> v_B \\
 E_{\text{total,A}} &> E_{\text{total,B}}
 \end{aligned}$$

Therefore, the ball has lost mechanical energy as it moves across the carpet. However, although the ball has lost mechanical energy, energy in the larger system has still been conserved. In this case, the missing energy is the work done by the carpet through applying a frictional force on the ball. In this case the carpet is doing negative work on the ball.

When an internal force does work on an object by an (for example, gravitational and spring forces), the total mechanical energy (KE + PE) of that object remains constant but the object's energy can change form. For example, as an object falls in a gravitational field from a high elevation to a lower elevation, some of the object's potential energy is changed into kinetic energy. However, the sum of the kinetic and potential energies remain constant. When the only forces doing work are internal forces, energy changes forms - from kinetic to potential (or vice versa); yet the total amount of mechanical energy is conserved.

8.3.2 Capacity to do Work

Energy is the capacity to do work. When positive work is done on an object, the system doing the work loses energy. In fact, **the energy lost by a system is exactly equal to the work**

done by the system. An object with larger potential energy has a greater capacity to do work.



Worked Example 49: Work Done on a System

Question: Show that a hammer of mass 2 kg does more work when dropped from a height of 10 m than when dropped from a height of 5 m. Confirm that the hammer has a greater potential energy at 10 m than at 5 m.

Answer

Step 5 : Determine what is given and what is required

We are given:

- the mass of the hammer, $m = 2$ kg
- height 1, $h_1 = 10$ m
- height 2, $h_2 = 5$ m

We are required to show that the hammer does more work when dropped from h_1 than from h_2 . We are also required to confirm that the hammer has a greater potential energy at 10 m than at 5 m.

Step 6 : Determine how to approach the problem

1. Calculate the work done by the hammer, W_1 , when dropped from h_1 using:

$$W_1 = F_g \cdot h_1.$$

2. Calculate the work done by the hammer, W_2 , when dropped from h_2 using:

$$W_2 = F_g \cdot h_2.$$

3. Compare W_1 and W_2
4. Calculate potential energy at h_1 and h_2 and compare using:

$$PE = m \cdot g \cdot h. \quad (8.2)$$

Step 7 : Calculate W_1

$$\begin{aligned} W_1 &= F_g \cdot h_1 \\ &= m \cdot g \cdot h_1 \\ &= (2 \text{ kg})(9.8 \text{ m} \cdot \text{s}^{-2})(10 \text{ m}) \\ &= 196 \text{ J} \end{aligned}$$

Step 8 : Calculate W_2

$$\begin{aligned} W_2 &= F_g \cdot h_2 \\ &= m \cdot g \cdot h_2 \\ &= (2 \text{ kg})(9.8 \text{ m} \cdot \text{s}^{-2})(5 \text{ m}) \\ &= 98 \text{ J} \end{aligned}$$

Step 9 : Compare W_1 and W_2

We have $W_1 = 196$ J and $W_2 = 98$ J. $W_1 > W_2$ as required.

Step 10 : Calculate potential energy

From 8.2, we see that:

$$\begin{aligned} PE &= m \cdot g \cdot h \\ &= F_g \cdot h \\ &= W \end{aligned}$$

This means that the potential energy is equal to the work done. Therefore, $PE_1 > PE_2$, because $W_1 > W_2$.

This leads us to the work-energy theorem.



Definition: Work-Energy Theorem

The work-energy theorem states that the work done on an object is equal to the change in its kinetic energy:

$$W = \Delta KE = KE_f - KE_i$$

The work-energy theorem is another example of the conservation of energy which you saw in Grade 10.



Worked Example 50: Work-Energy Theorem

Question: A brick of mass 1 kg is dropped from a height of 10 m. Calculate the work done on the brick at the point it hits the ground assuming that there is no air resistance?

Answer

Step 1 : Determine what is given and what is required

We are given:

- mass of the brick: $m=1$ kg
- initial height of the brick: $h_i=10$ m
- final height of the brick: $h_f=0$ m

We are required to determine the work done on the brick as it hits the ground.

Step 2 : Determine how to approach the problem

The brick is falling freely, so energy is conserved. We know that the work done is equal to the difference in kinetic energy. The brick has no kinetic energy at the moment it is dropped, because it is stationary. When the brick hits the ground, all the brick's potential energy is converted to kinetic energy.

Step 3 : Determine the brick's potential energy at h_i

$$\begin{aligned} PE &= m \cdot g \cdot h \\ &= (1 \text{ kg})(9,8 \text{ m} \cdot \text{s}^{-2})(10 \text{ m}) \\ &= 98 \text{ J} \end{aligned}$$

Step 4 : Determine the work done on the brick

The brick had 98 J of potential energy when it was released and 0 J of kinetic energy. When the brick hit the ground, it had 0 J of potential energy and 98 J of kinetic energy. Therefore $KE_i=0$ J and $KE_f=98$ J.

From the work-energy theorem:

$$\begin{aligned} W &= \Delta KE \\ &= KE_f - KE_i \\ &= 98 \text{ J} - 0 \text{ J} \\ &= 98 \text{ J} \end{aligned}$$

Step 5 : Write the final answer

98 J of work was done on the brick.

**Worked Example 51: Work-Energy Theorem 2**

Question: The driver of a 1 000 kg car traveling at a speed of $16,7 \text{ m}\cdot\text{s}^{-1}$ applies the car's brakes when he sees a red robot. The car's brakes provide a frictional force of 8000 N. Determine the stopping distance of the car.

Answer**Step 1 : Determine what is given and what is required**

We are given:

- mass of the car: $m=1\ 000 \text{ kg}$
- speed of the car: $v=16,7 \text{ m}\cdot\text{s}^{-1}$
- frictional force of brakes: $F=8\ 000 \text{ N}$

We are required to determine the stopping distance of the car.

Step 2 : Determine how to approach the problem

We apply the work-energy theorem. We know that all the car's kinetic energy is lost to friction. Therefore, the change in the car's kinetic energy is equal to the work done by the frictional force of the car's brakes.

Therefore, we first need to determine the car's kinetic energy at the moment of braking using:

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

This energy is equal to the work done by the brakes. We have the force applied by the brakes, and we can use:

$$W = F \cdot d$$

to determine the stopping distance.

Step 3 : Determine the kinetic energy of the car

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(1\ 000 \text{ kg})(16,7 \text{ m}\cdot\text{s}^{-1})^2 \\ &= 139\ 445 \text{ J} \end{aligned}$$

Step 4 : Determine the work doneAssume the stopping distance is d_0 . Then the work done is:

$$\begin{aligned} W &= F \cdot d \\ &= (-8\ 000 \text{ N})(d_0) \end{aligned}$$

The force has a negative sign because it acts in a direction opposite to the direction of motion.

Step 5 : Apply the work-energy theorem

The change in kinetic energy is equal to the work done.

$$\begin{aligned}\Delta KE &= W \\ KE_f - KE_i &= (-8\,000\text{ N})(d_0) \\ 0\text{ J} - 139\,445\text{ J} &= (-8\,000\text{ N})(d_0) \\ \therefore d_0 &= \frac{139\,445\text{ J}}{8\,000\text{ N}} \\ &= 17,4\text{ m}\end{aligned}$$

Step 6 : Write the final answer

The car stops in 17,4 m.



Important: A force only does work on an object for the time that it is in contact with the object. For example, a person pushing a trolley does work on the trolley, but the road does no work on the tyres of a car if they turn without slipping (the force is not applied over any distance because a different piece of tyre touches the road every instant).

Energy is conserved!

Important: Energy Conservation:

In the absence of friction, the work done on an object by a system is equal to the energy gained by the object.

$$\text{Work Done} = \text{Energy Transferred}$$

In the presence of friction, only some of the energy lost by the system is transferred to useful energy. The rest is lost to friction.

$$\text{Total Work Done} = \text{Useful Work Done} + \text{Work Done Against Friction}$$

In the example of a falling mass the potential energy is known as *gravitational potential energy* as it is the gravitational force exerted by the earth which causes the mass to accelerate towards the ground. The gravitational field of the earth is what does the work in this case.

Another example is a rubber-band. In order to stretch a rubber-band we have to do work on it. This means we transfer energy to the rubber-band and it gains potential energy. This potential energy is called *elastic potential energy*. Once released, the rubber-band begins to move and elastic potential energy is transferred into kinetic energy.

*Extension: Other forms of Potential Energy*

1. elastic potential energy - potential energy is stored in a compressed or extended spring or rubber band. This potential energy is calculated by:

$$\frac{1}{2}kx^2$$

where k is a constant that is a measure of the stiffness of the spring or rubber band and x is the extension of the spring or rubber band.

2. Chemical potential energy is related to the making and breaking of chemical bonds. For example, a battery converts chemical energy into electrical energy.
3. The electrical potential energy of an electrically charged object is defined as the work that must be done to move it from an infinite distance away to its present location, in the absence of any non-electrical forces on the object. This energy

is non-zero if there is another electrically charged object nearby otherwise it is given by:

$$k \frac{q_1 q_2}{d}$$

where k is Coulomb's constant. For example, an electric motor lifting an elevator converts electrical energy into gravitational potential energy.

4. Nuclear energy is the energy released when the nucleus of an atom is split or fused. A nuclear reactor converts nuclear energy into heat.

Some of these forms of energy will be studied in later chapters.

Activity :: Investigation : Energy Resources

Energy can be taken from almost anywhere. Power plants use many different types of energy sources, including oil, coal, nuclear, biomass (organic gases), wind, solar, geothermal (the heat from the earth's rocks is very hot underground and is used to turn water to steam), tidal and hydroelectric (waterfalls). Most power stations work by using steam to turn turbines which then drive generators and create an electric current.

Most of these sources are dependant upon the sun's energy, because without it we would not have weather for wind and tides. The sun is also responsible for growing plants which decompose into fossil fuels like oil and coal. All these sources can be put under 2 headings, renewable and non-renewable. Renewable sources are sources which will not run out, like solar energy and wind power. Non-renewable sources are ones which will run out eventually, like oil and coal.

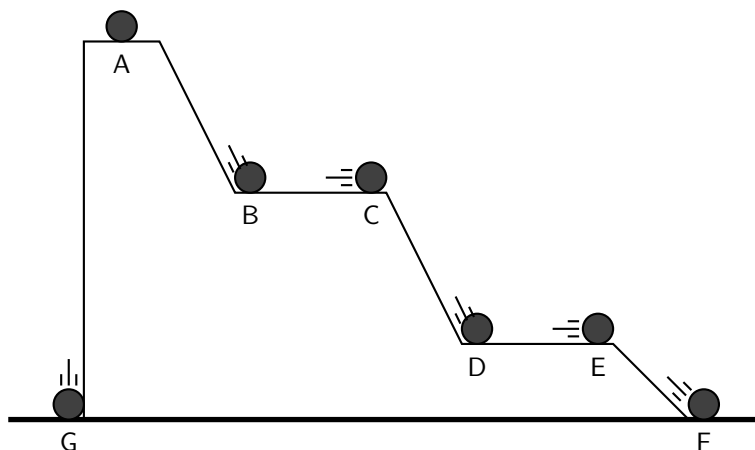
It is important that we learn to appreciate conservation in situations like this. The planet has a number of linked systems and if we don't appreciate the long-term consequences of our actions we run the risk of doing damage now that we will only suffer from in many years time.

Investigate two types of renewable and two types of non-renewable energy resources, listing advantages and disadvantages of each type. Write up the results as a short report.



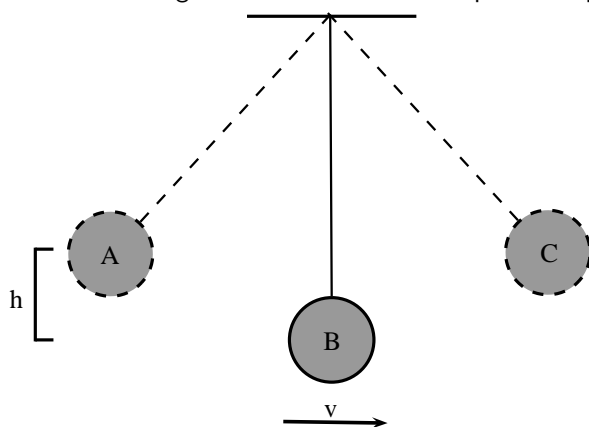
Exercise: Energy

1. Fill in the table with the missing information using the positions of the ball in the diagram below combined with the work-energy theorem.



position	KE	PE	v
A		50 J	
B		30 J	
C			
D		10 J	
E			
F			
G			

- A falling ball hits the ground at $10 \text{ m}\cdot\text{s}^{-1}$ in a vacuum. Would the speed of the ball be increased or decreased if air resistance were taken into account. Discuss using the work-energy theorem.
- A pendulum with mass 300g is attached to the ceiling. It is pulled up to point A which is a height $h = 30 \text{ cm}$ from the equilibrium position.



Calculate the speed of the pendulum when it reaches point B (the equilibrium point). Assume that there are no external forces acting on the pendulum.

8.4 Power

Now that we understand the relationship between work and energy, we are ready to look at a quantity that defines how long it takes for a certain amount of work to be done. For example, a mother pushing a trolley full of groceries can take 30 s or 60 s to push the trolley down an aisle. She does the same amount of work, but takes a different length of time. We use the idea of *power* to describe the rate at which work is done.



Definition: Power

Power is defined as the rate at which work is done or the rate at which energy is expended. The mathematical definition for power is:

$$P = F \cdot v \quad (8.3)$$

(8.3) is easily derived from the definition of work. We know that:

$$W = F \cdot d.$$

However, power is defined as the rate at which work is done. Therefore,

$$P = \frac{\Delta W}{\Delta t}.$$

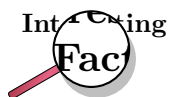
This can be written as:

$$\begin{aligned}
 P &= \frac{\Delta W}{\Delta t} \\
 &= \frac{\Delta(F \cdot d)}{\Delta t} \\
 &= F \frac{\Delta d}{\Delta t} \\
 &= F \cdot v
 \end{aligned}$$

The unit of power is watt (symbol W).

Activity :: Investigation : Watt

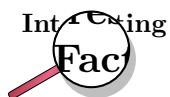
Show that the W is equivalent to $\text{J} \cdot \text{s}^{-1}$.



The unit watt is named after Scottish inventor and engineer James Watt (19 January 1736 - 19 August 1819) whose improvements to the steam engine were fundamental to the Industrial Revolution. A key feature of it was that it brought the engine out of the remote coal fields into factories.

Activity :: Research Project : James Watt

Write a short report 5 pages on the life of James Watt describing his many other inventions.



Historically, the *horsepower* (symbol hp) was the unit used to describe the power delivered by a machine. One horsepower is equivalent to approximately 750 W. The horsepower is sometimes used in the motor industry to describe the power output of an engine. Incidentally, the horsepower was derived by James Watt to give an indication of the power of his steam engine in terms of the power of a horse, which was what most people used to for example, turn a mill wheel.



Worked Example 52: Power Calculation 1

Question: Calculate the power required for a force of 10 N applied to move a 10 kg box at a speed of 1 ms over a frictionless surface.

Answer

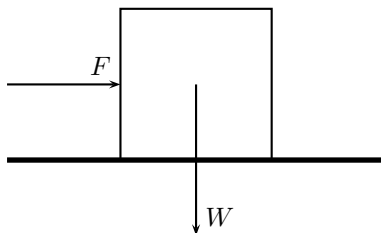
Step 1 : Determine what is given and what is required.

We are given:

- we are given the force, $F=10\text{ N}$
- we are given the speed, $v=1\text{ m}\cdot\text{s}^{-1}$

We are required to calculate the power required.

Step 2 : Draw a force diagram



Step 3 : Determine how to approach the problem

From the force diagram, we see that the weight of the box is acting at right angles to the direction of motion. The weight does not contribute to the work done and does not contribute to the power calculation.

We can therefore calculate power from:

$$P = F \cdot v$$

Step 4 : Calculate the power required

$$\begin{aligned} P &= F \cdot v \\ &= (10\text{ N})(1\text{ m} \cdot \text{s}^{-1}) \\ &= 10\text{ W} \end{aligned}$$

Step 5 : Write the final answer

10 W of power are required for a force of 10 N to move a 10 kg box at a speed of 1 ms over a frictionless surface.

Machines are designed and built to do work on objects. All machines usually have a power rating. The power rating indicates the rate at which that machine can do work upon other objects.

A car engine is an example of a machine which is given a power rating. The power rating relates to how rapidly the car can accelerate. Suppose that a 50 kW engine could accelerate the car from $0\text{ km} \cdot \text{hr}^{-1}$ to $60\text{ km} \cdot \text{hr}^{-1}$ in 16 s. Then a car with four times the power rating (i.e. 200 kW) could do the same amount of work in a quarter of the time. That is, a 200 kW engine could accelerate the same car from $0\text{ km} \cdot \text{hr}^{-1}$ to $60\text{ km} \cdot \text{hr}^{-1}$ in 4 s.

**Worked Example 53: Power Calculation 2**

Question: A forklift lifts a crate of mass 100 kg at a constant velocity to a height of 8 m over a time of 4 s. The forklift then holds the crate in place for 20 s. Calculate how much power the forklift exerts in lifting the crate? How much power does the forklift exert in holding the crate in place?

Answer

Step 1 : Determine what is given and what is required

We are given:

- mass of crate: $m=100$ kg
- height that crate is raised: $h=8$ m
- time taken to raise crate: $t_r=4$ s
- time that crate is held in place: $t_s=20$ s

We are required to calculate the power exerted.

Step 2 : Determine how to approach the problem

We can use:

$$P = F \frac{\Delta x}{\Delta t}$$

to calculate power. The force required to raise the crate is equal to the weight of the crate.

Step 3 : Calculate the power required to raise the crate

$$\begin{aligned} P &= F \frac{\Delta x}{\Delta t} \\ &= m \cdot g \frac{\Delta x}{\Delta t} \\ &= (100 \text{ kg})(9,8 \text{ m} \cdot \text{s}^{-2}) \frac{8 \text{ m}}{4 \text{ s}} \\ &= 1\,960 \text{ W} \end{aligned}$$

Step 4 : Calculate the power required to hold the crate in place

While the crate is being held in place, there is no displacement. This means there is no work done on the crate and therefore there is no power exerted.

Step 5 : Write the final answer

1 960 W of power is exerted to raise the crate and no power is exerted to hold the crate in place.

Activity :: Experiment : Simple measurements of human power

You can perform various physical activities, for example lifting measured weights or climbing a flight of stairs to estimate your output power, using a stop watch.

Note: the human body is not very efficient in these activities, so your actual power will be much greater than estimated here.

**Exercise: Power**

1. [IEB 2005/11 HG] Which of the following is equivalent to the SI unit of power:

- (a) $\text{V} \cdot \text{A}$
- (b) $\text{V} \cdot \text{A}^{-1}$
- (c) $\text{kg} \cdot \text{m} \cdot \text{s}^{-1}$
- (d) $\text{kg} \cdot \text{m} \cdot \text{s}^{-2}$

2. Two students, Bill and Bob, are in the weight lifting room of their local gym. Bill lifts the 50 kg barbell over his head 10 times in one minute while Bob lifts the 50 kg barbell over his head 10 times in 10 seconds. Who does the most work? Who delivers the most power? Explain your answers.
3. Jack and Jill ran up the hill. Jack is twice as massive as Jill; yet Jill ascended the same distance in half the time. Who did the most work? Who delivered the most power? Explain your answers.
4. Alex (mass 60 kg) is training for the Comrades Marathon. Part of Alex's training schedule involves push-ups. Alex does his push-ups by applying a force to elevate his center-of-mass by 20 cm. Determine the number of push-ups that Alex must do in order to do 10 J of work. If Alex does all this work in 60 s, then determine Alex's power.
5. When doing a chin-up, a physics student lifts her 40 kg body a distance of 0.25 m in 2 s. What is the power delivered by the student's biceps?
6. The unit of power that is used on a monthly electricity account is *kilowatt-hours* (symbol kWh). This is a unit of energy delivered by the flow of 1 kW of electricity for 1 hour. Show how many joules of energy you get when you buy 1 kWh of electricity.
7. An escalator is used to move 20 passengers every minute from the first floor of a shopping mall to the second. The second floor is located 5-meters above the first floor. The average passenger's mass is 70 kg. Determine the power requirement of the escalator in order to move this number of passengers in this amount of time.
8. Calculate the power required for an electric motor to pump 20kg of water up to ground level from a borehole of depth 10m in half a minute.

8.5 Important Equations and Quantities

Units					
Quantity	Symbol	Unit	S.I. Units		Direction
velocity	\vec{v}	—	$\frac{m}{s}$	or $m.s^{-1}$	✓
momentum	\vec{p}	—	$\frac{kg.m}{s}$	or $kg.m.s^{-1}$	✓
energy	E	J	$\frac{kg.m^2}{s^2}$	or $kg.m^2.s^{-2}$	—
Work	W	J	$N.m$	or $kg.m^2.s^{-2}$	—
Kinetic Energy	E_K	J	$N.m$	or $kg.m^2.s^{-2}$	—
Potential Energy	E_P	J	$N.m$	or $kg.m^2.s^{-2}$	—
Mechanical Energy	U	J	$N.m$	or $kg.m^2.s^{-2}$	—
Power	P	W	$N.m.s^{-1}$	or $kg.m^2.s^{-3}$	—

Table 8.1: Commonly used units

Momentum:

$$\vec{p} = m\vec{v} \quad (8.4)$$

Kinetic energy:

$$E_k = \frac{1}{2}m\vec{v}^2 \quad (8.5)$$

Principle of Conservation of Energy: Energy is never created nor destroyed, but is merely transformed from one form to another.

Conservation of Mechanical Energy: In the absence of friction, the total mechanical energy of an object is conserved.

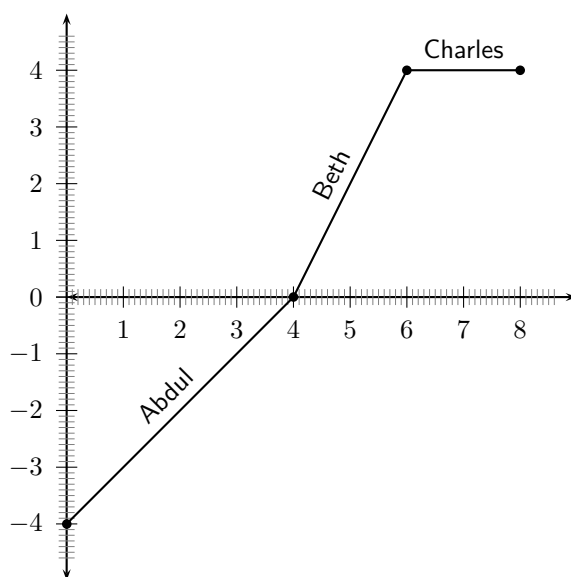
When a force moves in the direction along which it acts, work is done.

Work is the process of converting energy.

Energy is the ability to do work.

8.6 End of Chapter Exercises

- The force vs. displacement graph shows the amount of force applied to an object by three different people. Abdul applies force to the object for the first 4 m of its displacement, Beth applies force from the 4 m point to the 6 m point, and Charles applies force from the 6 m point to the 8 m point. Calculate the work done by each person on the object? Which of the three does the most work on the object?

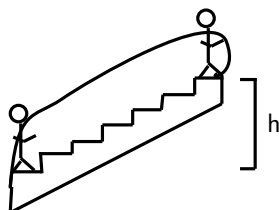


- How much work does a person do in pushing a shopping trolley with a force of 200 N over a distance of 80 m in the direction of the force?
- How much work does the force of gravity do in pulling a 20 kg box down a 45° frictionless inclined plane of length 18 m?
- [IEB 2001/11 HG1] Of which one of the following quantities is $\text{kg}\cdot\text{m}^2\cdot\text{s}^{-3}$ the base S.I. unit?
 - Energy
 - Force
 - Power
 - Momentum
- [IEB 2003/11 HG1] A motor is used to raise a mass m through a vertical height h in time t .
What is the power of the motor while doing this?
 - $mght$
 - $\frac{mgh}{t}$
 - $\frac{mgt}{h}$
 - $\frac{ht}{mg}$
- [IEB 2002/11 HG1] An electric motor lifts a load of mass M vertically through a height h at a constant speed v . Which of the following expressions can be used to correctly calculate the power transferred by the motor to the load while it is lifted at a constant speed?

- (a) Mgh
- (b) $Mgh + \frac{1}{2}Mv^2$
- (c) Mgv
- (d) $Mgv + \frac{1}{2} \frac{Mv^3}{h}$

7. [IEB 2001/11 HG1] An escalator is a moving staircase that is powered by an electric motor. People are lifted up the escalator at a constant speed of v through a vertical height h .

What is the energy gained by a person of mass m standing on the escalator when he is lifted from the bottom to the top?



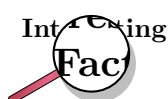
- (a) mgh
 - (b) $mgh \sin \theta$
 - (c) $\frac{mgh}{\sin \theta}$
 - (d) $\frac{1}{2}mv^2$
8. [IEB 2003/11 HG1] In which of the following situations is there no work done on the object?
- (a) An apple falls to the ground.
 - (b) A brick is lifted from the ground to the top of a building.
 - (c) A car slows down to a stop.
 - (d) A box moves at constant velocity across a frictionless horizontal surface.

Chapter 9

Doppler Effect - Grade 12

9.1 Introduction

Have you noticed how the pitch of a police car siren changes as the car passes by or how the pitch of a radio box on the pavement changes as you drive by? This effect is known as the **Doppler Effect** and will be studied in this chapter.



The Doppler Effect is named after Johann Christian Andreas Doppler (29 November 1803 - 17 March 1853), an Austrian mathematician and physicist who first explained the phenomenon in 1842.

9.2 The Doppler Effect with Sound and Ultrasound

As seen in the introduction, there are two situations which lead to the Doppler Effect:

1. When the source moves relative to the observer, for example the pitch of a car hooter as it passes by.
2. When the observer moves relative to the source, for example the pitch of a radio on the pavement as you drive by.



Definition: Doppler Effect

The Doppler effect is the apparent change in frequency and wavelength of a wave when the observer and the source of the wave move relative to each other.

We experience the Doppler effect quite often in our lives, without realising that it is science taking place. The changing sound of a taxi hooter or ambulance as it drives past are examples of this as you have seen in the introduction.

The question is how does the Doppler effect take place. Let us consider a source of sound waves with a constant frequency and amplitude. The sound waves can be drawn as concentric circles where each circle represents another wavefront, like in figure 9.1 below.

The sound source is the dot in the middle and is stationary. For the Doppler effect to take place, the source must be moving *relative* to the observer. Let's consider the following situation: The source (dot) emits one peak (represented by a circle) that moves away from the source at the same rate in all directions. The distance between the peaks represents the wavelength of the sound. The closer together the peaks, the higher the frequency (or pitch) of the sound.

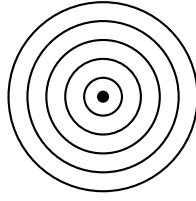
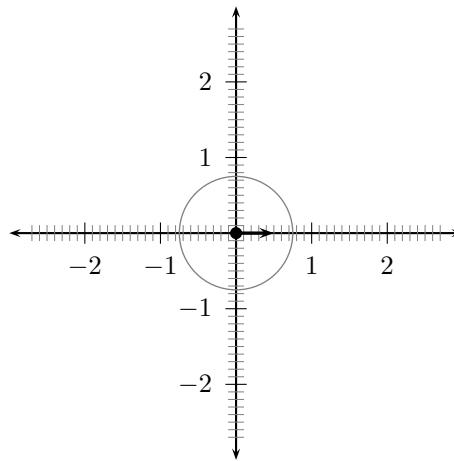
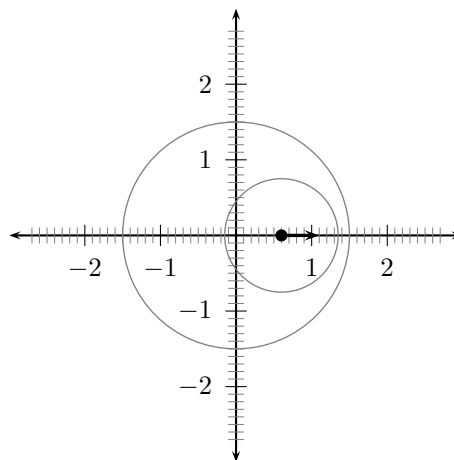


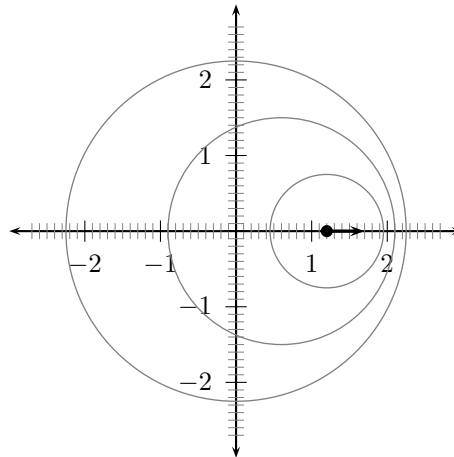
Figure 9.1: Stationary sound source



As this peak moves away, the source also moves and then emits the second peak. Now the two circles are not concentric any more, but on the one side they are closer together and on the other side they are further apart. This is shown in the next diagram.



If the source continues moving at the same speed in the same direction (i.e. with the same velocity which you will learn more about later), then the distance between peaks on the right of the source is constant. The distance between peaks on the left is also constant but they are different on the left and right.



This means that the time between peaks on the right is less so the frequency is higher. It is higher than on the left and higher than if the source were not moving at all.

On the left hand side the peaks are further apart than on the right and further apart than if the source were at rest - this means the frequency is lower.

When a car approaches you, the sound waves that reach you have a shorter wavelength and a higher frequency. You hear a higher sound. When the car moves away from you, the sound waves that reach you have a longer wavelength and lower frequency. You hear a lower sound.

This change in frequency can be calculated by using:

$$f_L = \frac{v + v_L}{v + v_S} f_S \quad (9.1)$$

where f_L is the frequency perceived by the listener,
 f_S is the frequency of the source,
 v is the velocity of the waves,
 v_L the velocity of the listener and
 v_S the velocity of the source.

Note: Velocity is a vector and has magnitude and direction. It is very important to get the signs of the velocities correct here:

Source moves towards listener	v_S : negative
Source moves away from listener	v_S : positive
Listener moves towards source	v_L : positive
Listener moves away from source	v_L : negative



Worked Example 54: The Doppler Effect for Sound

Question: The siren of an ambulance has a frequency of 700 Hz. You are standing on the pavement. If the ambulance drives past you at a speed of $20 \text{ m}\cdot\text{s}^{-1}$, what frequency will you hear, when

- the ambulance is approaching you
- the ambulance is driving away from you

Take the speed of sound to be $340 \text{ m}\cdot\text{s}^{-1}$.

Answer

Step 1 : Determine how to approach the problem based on what is given

$$f_L = \frac{v + v_L}{v + v_S} f_S$$

$$\begin{aligned}
 f_s &= 700\text{Hz} \\
 v &= 340\text{ m} \cdot \text{s}^{-1} \\
 v_L &= 0 \\
 v_S &= -20\text{ m} \cdot \text{s}^{-1} \text{ for (a) and} \\
 v_S &= +20\text{ m} \cdot \text{s}^{-1} \text{ for (b)}
 \end{aligned}$$

Step 2 : Determine f_L when ambulance is approaching

$$\begin{aligned}
 f_L &= \frac{340 + 0}{340 - 20}(700) \\
 &= 743,75\text{Hz}
 \end{aligned}$$

Step 3 : Determine f_L when ambulance has passed

$$\begin{aligned}
 f_L &= \frac{340 + 0}{340 + 20}(700) \\
 &= 661,11\text{Hz}
 \end{aligned}$$



Worked Example 55: The Doppler Effect for Sound 2

Question: What is the frequency heard by a person driving at $15\text{ m} \cdot \text{s}^{-1}$ toward a factory whistle that is blowing at a frequency of 800 Hz . Assume that the speed of sound is $340\text{ m} \cdot \text{s}^{-1}$.

Answer

Step 1 : Determine how to approach the problem based on what is given

We can use

$$f_L = \frac{v + v_L}{v + v_S} f_S$$

with:

$$\begin{aligned}
 v &= 340\text{ m} \cdot \text{s}^{-1} \\
 v_L &= +15\text{ m} \cdot \text{s}^{-1} \\
 v_S &= 0\text{ m} \cdot \text{s}^{-1} \\
 f_S &= 800\text{ Hz} \\
 f_L &= ?
 \end{aligned}$$

The listener is moving towards the source, so v_L is positive.

Step 2 : Calculate the frequency

$$\begin{aligned}
 f_L &= \frac{v + v_L}{v + v_S} f_S \\
 &= \frac{340,6\text{ m} \cdot \text{s}^{-1} + 15\text{ m} \cdot \text{s}^{-1}}{340,6\text{ m} \cdot \text{s}^{-1} + 0\text{ m} \cdot \text{s}^{-1}} (800\text{ Hz}) \\
 &= 835\text{ Hz}
 \end{aligned}$$

Step 3 : Write the final answer

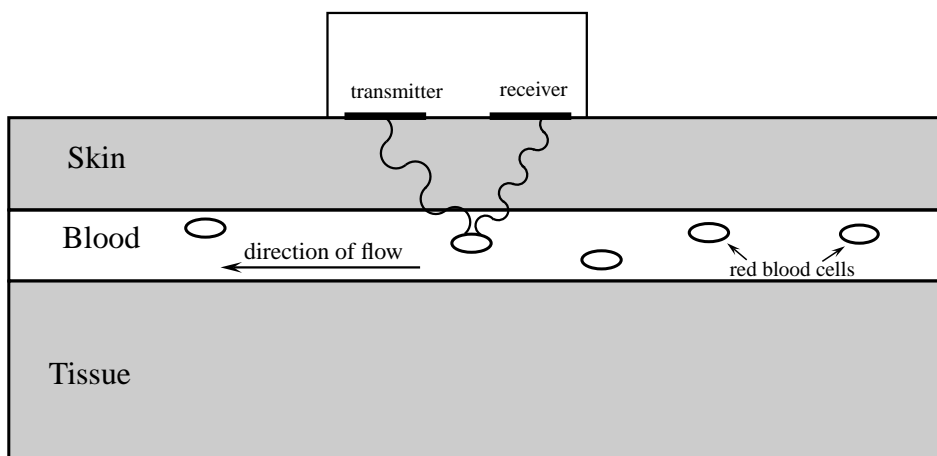
The driver hears a frequency of 835 Hz .



Radar-based speed-traps use the Doppler Effect. The radar gun emits radio waves of a specific frequency. When the car is standing still, the waves reflected waves are the same frequency as the waves emitted by the radar gun. When the car is moving the Doppler frequency shift can be used to determine the speed of the car.

9.2.1 Ultrasound and the Doppler Effect

Ultrasonic waves (ultrasound) are sound waves with a frequency greater than 20 000 Hz (the upper limit of human hearing). These waves can be used in medicine to determine the direction of blood flow. The device, called a Doppler flow meter, sends out sound waves. The sound waves can travel through skin and tissue and will be reflected by moving objects in the body (like blood). The reflected waves return to the flow meter where its frequency (received frequency) is compared to the transmitted frequency. Because of the Doppler effect, blood that is moving towards the flow meter will change the sound to a higher frequency and blood that is moving away from the flow meter will cause a lower frequency.



Ultrasound can be used to determine whether blood is flowing in the right direction in the circulation system of unborn babies, or identify areas in the body where blood flow is restricted due to narrow veins. The use of ultrasound equipment in medicine is called sonography or ultrasonography.



Exercise: The Doppler Effect with Sound

1. Suppose a train is approaching you as you stand on the platform at the station. As the train approaches the station, it slows down. All the while, the engineer is sounding the hooter at a constant frequency of 400 Hz. Describe the pitch and the changes in pitch that you hear.
2. Passengers on a train hear its whistle at a frequency of 740 Hz. Anja is standing next to the train tracks. What frequency does Anja hear as the train moves directly toward her at a speed of $25 \text{ m}\cdot\text{s}^{-1}$?
3. A small plane is taxiing directly away from you down a runway. The noise of the engine, as the pilot hears it, has a frequency 1,15 times the frequency that you hear. What is the speed of the plane?

9.3 The Doppler Effect with Light

Light is a wave and earlier you learnt how you can study the properties of one wave and apply the same ideas to another wave. The same applies to sound and light. We know the Doppler effect affects sound waves when the source is moving. Therefore, if we apply the Doppler effect to light, the frequency of the emitted light should change when the source of the light is moving relative to the observer.

When the frequency of a sound wave changes, the sound you hear changes. When the frequency of light changes, the colour you would see changes.

This means that the Doppler effect can be observed by a change in sound (for sound waves) and a change in colour (for light waves). Keep in mind that there are sounds that we cannot hear (for example ultrasound) and light that we cannot see (for example ultraviolet light).

We can apply all the ideas that we learnt about the Doppler effect to light. When talking about light we use slightly different names to describe what happens. If you look at the colour spectrum (more details Chapter 15) then you will see that blue light has shorter wavelengths than red light. If you are in the middle of the visible colours then longer wavelengths are more red and shorter wavelengths are more blue. So we call shifts towards longer wavelengths "red-shifts" and shifts towards shorter wavelengths "blue-shifts".

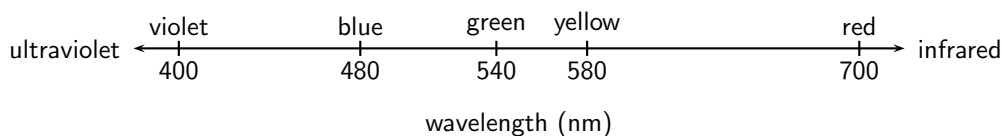


Figure 9.2: Blue light has shorter wavelengths than red light.

A shift in wavelength is the same as a shift in frequency. Longer wavelengths of light have lower frequencies and shorter wavelengths have higher frequencies. From the Doppler effect we know that when things move towards you any waves they emit that you measure are shifted to shorter wavelengths (blueshifted). If things move away from you, the shift is to longer wavelengths (redshifted).

9.3.1 The Expanding Universe

Stars emit light, which is why we can see them at night. Galaxies are huge collections of stars. An example is our own Galaxy, the Milky Way, of which our sun is only one of the millions of stars! Using large telescopes like the Southern African Large Telescope (SALT) in the Karoo, astronomers can measure the light from distant galaxies. The spectrum of light can tell us what elements are in the stars in the galaxies because each element emits/absorbs light at particular wavelengths (called spectral lines). If these lines are observed to be shifted from their usual wavelengths to shorter wavelengths, then the light from the galaxy is said to be *blueshifted*. If the spectral lines are shifted to longer wavelengths, then the light from the galaxy is said to be *redshifted*. If we think of the blueshift and redshift in Doppler effect terms, then a blueshifted galaxy would appear to be moving *towards* us (the observers) and a redshifted galaxy would appear to be moving *away* from us.

Important:

- If the light source is moving away from the observer (positive velocity) then the observed frequency is lower and the observed wavelength is greater (redshifted).
- If the source is moving towards (negative velocity) the observer, the observed frequency is higher and the wavelength is shorter (blueshifted).

Edwin Hubble (20 November 1889 - 28 September 1953) measured the Doppler shift of a large sample of galaxies. He found that the light from distant galaxies is *redshifted* and he discovered that there is a proportionality relationship between the *redshift* and the *distance* to the galaxy.

Galaxies that are further away always appear more redshifted than nearby galaxies. Remember that a redshift in Doppler terms means a velocity of the light source *away* from the observer. So why do all distant galaxies appear to be moving away from our Galaxy?

The reason is that the universe is expanding! The galaxies are not actually moving themselves, rather the *space* between them is expanding!

9.4 Summary

1. The Doppler Effect is the apparent change in frequency and wavelength of a wave when the observer and source of the wave move relative to each other.
2. The following equation can be used to calculate the frequency of the wave according to the observer or listener:

$$f_L = \frac{v + v_L}{v + v_S} f_S$$

3. If the direction of the wave from the listener to the source is chosen as positive, the velocities have the following signs:

Source moves towards listener	v_S : negative
Source moves away from listener	v_S : positive
Listener moves towards source	v_L : positive
Listener moves away from source	v_L : negative

4. The Doppler Effect can be observed in all types of waves, including ultrasound, light and radiowaves.
5. Sonography makes use of ultrasound and the Doppler Effect to determine the direction of blood flow.
6. Light is emitted by stars. Due to the Doppler Effect, the frequency of this light decreases and the stars appear red. This is called a red shift and means that the stars are moving away from the Earth. This means that the Universe is expanding.

9.5 End of Chapter Exercises

1. Write a definition for each of the following terms.
 - (a) Doppler Effect
 - (b) Red-shift
 - (c) Ultrasound
2. Explain how the Doppler Effect is used to determine the direction of blood flow in veins.
3. The hooter of an approaching taxi has a frequency of 500 Hz. If the taxi is travelling at $30 \text{ m}\cdot\text{s}^{-1}$ and the speed of sound is $300 \text{ m}\cdot\text{s}^{-1}$, calculate the frequency of sound that you hear when
 - (a) the taxi is approaching you.
 - (b) the taxi passed you and is driving away.
4. A truck approaches you at an unknown speed. The sound of the trucks engine has a frequency of 210 Hz, however you hear a frequency of 220 Hz. The speed of sound is $340 \text{ m}\cdot\text{s}^{-1}$.
 - (a) Calculate the speed of the truck.
 - (b) How will the sound change as the truck passes you? Explain this phenomenon in terms of the wavelength and frequency of the sound.

5. A police car is driving towards a fleeing suspect at $\frac{v}{35}$ m.s⁻¹, where v is the speed of sound. The frequency of the police car's siren is 400 Hz. The suspect is running away at $\frac{v}{68}$. What frequency does the suspect hear?
6. (a) Why are ultrasound waves used in sonography and not sound waves?
(b) Explain how the Doppler effect is used to determine the direction of flow of blood in veins.

Chapter 10

Colour - Grade 12

10.1 Introduction

The light that human beings can see is called *visible light*. Visible light is actually just a small part of the large spectrum of electromagnetic radiation which you will learn more about in Chapter 15. We can think of electromagnetic radiation and visible light as transverse waves. We know that transverse waves can be described by their amplitude, frequency (or wavelength) and velocity. The velocity of a wave is given by the product of its frequency and wavelength:

$$v = f \times \lambda \quad (10.1)$$

However, electromagnetic radiation, including visible light, is special because, no matter what the frequency, it all moves at a **constant velocity** (in vacuum) which is known as the speed of light. The speed of light has the symbol c and is:

$$c = 3 \times 10^8 \text{ m.s}^{-1}$$

Since the *speed of light* is c , we can then say:

$$c = f \times \lambda \quad (10.2)$$

10.2 Colour and Light

Our eyes are sensitive to visible light over a range of wavelengths from 390 nm to 780 nm ($1 \text{ nm} = 1 \times 10^{-9} \text{ m}$). The different **colours** of light we see are related to specific *frequencies* (and *wavelengths*) of visible light. The wavelengths and frequencies are listed in table 10.1.

Colour	Wavelength range (nm)	Frequency range (Hz)
violet	390 - 455	$769 - 659 \times 10^{12}$
blue	455 - 492	$659 - 610 \times 10^{12}$
green	492 - 577	$610 - 520 \times 10^{12}$
yellow	577 - 597	$520 - 503 \times 10^{12}$
orange	597 - 622	$503 - 482 \times 10^{12}$
red	622 - 780	$482 - 385 \times 10^{12}$

Table 10.1: Colours, wavelengths and frequencies of light in the visible spectrum.

You can see from table 10.1 that **violet** light has the *shortest wavelengths* and *highest frequencies* while **red** light has the *longest wavelengths* and *lowest frequencies*.



Worked Example 56: Calculating the frequency of light given the wavelength**Question:** A streetlight emits light with a wavelength of 520 nm.

1. What colour is the light? (Use table 10.1 to determine the colour)
2. What is the frequency of the light?

Answer**Step 1 : What is being asked and what information are we given?**

We need to determine the colour and frequency of light with a wavelength of $\lambda = 520 \text{ nm} = 520 \times 10^{-9} \text{ m}$.

Step 2 : Compare the wavelength of the light to those given in table 10.1

We see from table 10.1 that light with wavelengths between 492 - 577 nm is green. 520 nm falls into this range, therefore the colour of the light is green.

Step 3 : Next we need to calculate the frequency of the light

We know that

$$c = f \times \lambda$$

We know c and we are given that $\lambda = 520 \times 10^{-9} \text{ m}$. So we can substitute in these values and solve for the frequency f . (**NOTE:** Don't forget to always change units into S.I. units! $1 \text{ nm} = 1 \times 10^{-9} \text{ m}$.)

$$\begin{aligned} f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^8}{520 \times 10^{-9}} \\ &= 577 \times 10^{12} \text{ Hz} \end{aligned}$$

The frequency of the green light is $577 \times 10^{12} \text{ Hz}$

**Worked Example 57: Calculating the wavelength of light given the frequency****Question:** A streetlight also emits light with a frequency of $490 \times 10^{12} \text{ Hz}$.

1. What colour is the light? (Use table 10.1 to determine the colour)
2. What is the wavelength of the light?

Answer**Step 1 : What is being asked and what information are we given?**

We need to find the colour and wavelength of light which has a frequency of $490 \times 10^{12} \text{ Hz}$ and which is emitted by the streetlight.

Step 2 : Compare the wavelength of the light to those given in table 10.1

We can see from table 10.1 that orange light has frequencies between $503 - 482 \times 10^{12} \text{ Hz}$. The light from the streetlight has $f = 490 \times 10^{12} \text{ Hz}$ which fits into this range. Therefore the light must be orange in colour.

Step 3 : Next we need to calculate the wavelength of the light

We know that

$$c = f \times \lambda$$

We know $c = 3 \times 10^8 \text{ m}\cdot\text{s}^{-1}$ and we are given that $f = 490 \times 10^{12} \text{ Hz}$. So we can substitute in these values and solve for the wavelength λ .

$$\begin{aligned}\lambda &= \frac{c}{f} \\ &= \frac{3 \times 10^8}{490 \times 10^{12}} \\ &= 6.122 \times 10^{-7} \text{ m} \\ &= 612 \times 10^{-9} \text{ m} \\ &= 612 \text{ nm}\end{aligned}$$

Therefore the orange light has a wavelength of 612 nm.



Worked Example 58: Frequency of Green

Question: The wavelength of green light ranges between 500 nm and 565 nm. Calculate the range of frequencies that correspond to this range of wavelengths.

Answer

Step 1 : Determine how to approach the problem

Use

$$c = f \times \lambda$$

to determine f .

Step 2 : Calculate frequency corresponding to upper limit of wavelength range

$$\begin{aligned}c &= f \times \lambda \\ f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^8 \text{ m}\cdot\text{s}^{-1}}{565 \times 10^{-9} \text{ m}} \\ &= 5,31 \times 10^{14} \text{ Hz}\end{aligned}$$

Step 3 : Calculate frequency corresponding to lower limit of wavelength range

$$\begin{aligned}c &= f \times \lambda \\ f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^8 \text{ m}\cdot\text{s}^{-1}}{500 \times 10^{-9} \text{ m}} \\ &= 6,00 \times 10^{14} \text{ Hz}\end{aligned}$$

Step 4 : Write final answer

The range of frequencies of green light is $5,31 \times 10^{14} \text{ Hz}$ to $6,00 \times 10^{14} \text{ Hz}$.



Exercise: Calculating wavelengths and frequencies of light

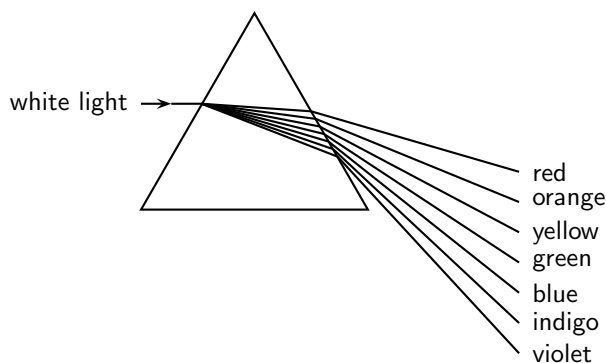
1. Calculate the frequency of light which has a wavelength of 400 nm. (Remember to use S.I. units)
2. Calculate the wavelength of light which has a frequency of 550×10^{12} Hz.
3. What colour is light which has a wavelength of 470×10^{-9} m and what is its frequency?
4. What is the wavelength of light with a frequency of 510×10^{12} Hz and what is its color?

10.2.1 Dispersion of white light

White light, like the light which comes from the sun, is made up of *all* the visible wavelengths of light. In other words, white light is a *combination* of all the colours of visible light.

You learnt that the speed of light is different in different substances. The speed of light in different substances depends on the frequency of the light. For example, when white light travels through glass, light of the different frequencies is slowed down by different amounts. The lower the frequency, the less the speed is reduced which means that red light (lowest frequency) is slowed down *less* than violet light (highest frequency). We can see this when white light is incident on a glass prism.

Have a look at the picture below. When the white light hits the edge of the prism, the light which travels through the glass is refracted as it moves from the less dense medium (air) to the more dense medium (glass).



- The **red** light which is slowed down the *least*, is refracted the *least*.
- The **violet** light which is slowed down the *most*, is refracted the *most*.

When the light hits the other side of the prism it is again refracted but the angle of the prism edge allows the light to remain separated into its different colours. White light is therefore separated into its different colours by the prism and we say that the white light has been **dispersed** by the prism.

The dispersion effect is also responsible for why we see rainbows. When sunlight hits drops of water in the atmosphere, the white light is dispersed into its different colours by the water.

10.3 Addition and Subtraction of Light

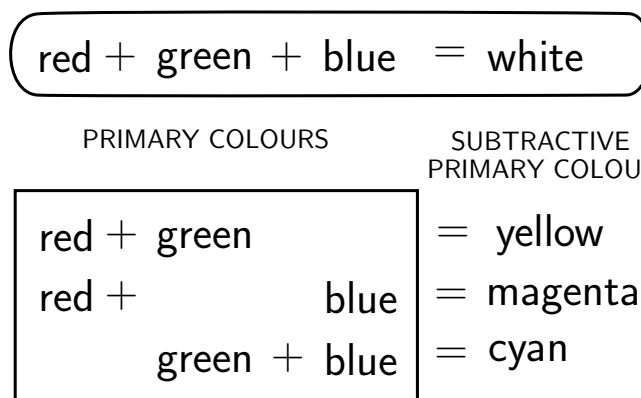
10.3.1 Additive Primary Colours

The primary colours of light are **red**, **green** and **blue**. When all the primary colours are superposed (added together), white light is produced. Red, green and blue are therefore called the

additive primary colours. All the other colours can be produced by different combinations of red, green and blue.

10.3.2 Subtractive Primary Colours

The subtractive primary colours are obtained by subtracting one of the three additive primary colours from white light. The subtractive primary colours are **yellow**, **magenta** and **cyan**. Magenta appears as a pinkish-purplish colour and cyan looks greenish-blue. You can see how the primary colours of light add up to the different subtractive colours in the illustration below.



Activity :: Experiment : Colours of light

Aim:

To investigate the additive properties of colours and determine the complementary colours of light.

Apparatus:

You will need two battery operated torches with flat bulb fronts, a large piece of white paper, and some pieces of cellophane paper of the following colours: red, blue, green, yellow, cyan, magenta. (You should easily be able to get these from a newsagents.)

Make a table in your workbook like the one below:

Colour 1	Colour 2	Final colour prediction	Final colour measured
red	blue		
red	green		
green	blue		
magenta	green		
yellow	blue		
cyan	red		

Before you begin your experiment, use what you know about colours of light to write down in the third column "Final colour prediction", what you think the result of adding the two colours of light will be. You will then be able to test your predictions by making the following measurements:

Method:

Proceed according to the table above. Put the correct colour of cellophane paper over each torch bulb. e.g. the first test will be to put red cellophane on one torch and blue cellophane on the other. Switch on the torch with the red cellophane over it and shine it onto the piece of white paper.

What colour is the light?

Turn off that torch and turn on the one with blue cellophane and shine it onto the white paper.

What colour is the light?

Now shine both torches with their cellophane coverings onto the same spot on the white paper. What is the colour of the light produced? Write this down in the fourth column of your table.

Repeat the experiment for the other colours of cellophane so that you can complete your table.

Questions:

1. How did your predictions match up to your measurements?
 2. Complementary colours of light are defined as the colours of light which, when added to one of the primary colours, produce white light. From your completed table, write down the complementary colours for red, blue and green.
-

10.3.3 Complementary Colours

Complementary colours are two colours of light which add together to give white.

Activity :: Investigation : Complementary colours for red, green and blue

Complementary colours are two colours which add together to give white. Place a tick in the box where the colours in the first column added to the colours in the top row give white.

	magenta (=red+blue)	yellow (=red+green)	cyan (=blue+green)
red			
green			
blue			

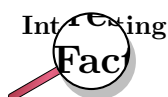
You should have found that the complementary colours for red, green and blue are:

- Red and Cyan
- Green and Magenta
- Blue and Yellow

10.3.4 Perception of Colour

The light-sensitive lining on the back inside half of the human eye is called the retina. The retina contains two kinds of light sensitive cells or *photoreceptors*: the rod cells (sensitive to low light) and the cone cells (sensitive to normal daylight) which enable us to see. The rods are not sensitive to colour but work well in dimly lit conditions. This is why it is possible to see in a dark room, but it is hard to see any colours. Only your rods are sensitive to the low light levels and so you can only see in black, white and grey. The cones enable us to see colours. Normally, there are three kinds of cones, each containing a different pigment. The cones are activated when the pigments absorb light. The three types of cones are sensitive to (i.e. absorb) red, blue and green light respectively. Therefore we can perceive *all* the different colours in the visible spectrum when the different types of cones are stimulated by different amounts since they are just combinations of the three primary colours of light.

The rods and cones have different response times to light. The cones react quickly when bright light falls on them. The rods take a longer time to react. This is why it takes a while (about 10 minutes) for your eyes to adjust when you enter a dark room after being outside on a sunny day.



Color blindness in humans is the inability to perceive differences between some or all colors that other people can see. Most often it is a genetic problem, but may also occur because of eye, nerve, or brain damage, or due to exposure to certain chemicals. The most common forms of human color blindness result from problems with either the middle or long wavelength sensitive cone systems, and involve difficulties in discriminating reds, yellows, and greens from one another. This is called "red-green color blindness". Other forms of color blindness are much rarer. They include problems in discriminating blues from yellows, and the rarest forms of all, complete color blindness or monochromasy, where one cannot distinguish any color from grey, as in a black-and-white movie or photograph.



Worked Example 59: Seeing Colours

Question: When blue and green light fall on an eye, is cyan light being created? Discuss.

Answer

Cyan light is not created when blue and green light fall on the eye. The blue and green receptors are stimulated to make the brain believe that cyan light is being created.

10.3.5 Colours on a Television Screen

If you look very closely at a colour cathode-ray television screen or cathode-ray computer screen, you will see that there are very many small red, green and blue dots called *phosphors* on it. These dots are caused to fluoresce (glow brightly) when a beam of electrons from the cathode-ray tube behind the screen hits them. Since different combinations of the three primary colours of light can produce any other colour, only red, green and blue dots are needed to make pictures containing all the colours of the visible spectrum.



Exercise: Colours of light

1. List the three primary colours of light.
 2. What is the term for the phenomenon whereby white light is split up into its different colours by a prism?
 3. What is meant by the term "complementary colour" of light?
 4. When white light strikes a prism which colour of light is refracted the most and which is refracted the least? Explain your answer in terms of the speed of light in a medium.
-

10.4 Pigments and Paints

We have learnt that white light is a combination of all the colours of the visible spectrum and that each colour of light is related to a different frequency. But what gives everyday objects around us their different colours?

Pigments are substances which give an object its colour by absorbing certain frequencies of light and reflecting other frequencies. For example, a red pigment absorbs all colours of light except red which it reflects. Paints and inks contain pigments which gives the paints and inks different colours.

10.4.1 Colour of opaque objects

Objects which you *cannot* see through (i.e. they are not transparent) are called **opaque**. Examples of some opaque objects are metals, wood and bricks. The colour of an opaque object is determined by the colours (therefore *frequencies*) of light which it *reflects*. For example, when white light strikes a blue opaque object such as a ruler, the ruler will absorb all frequencies of light *except* blue, which will be reflected. The reflected blue light is the light which makes it into our eyes and therefore the object will appear blue.

Opaque objects which appear white do not absorb any light. They reflect all the frequencies. Black opaque objects absorb all frequencies of light. They do not reflect at all and therefore appear to have no colour.



Worked Example 60: Colour of Opaque Objects

Question: If we shine white light on a sheet of paper that can only reflect green light, what is the colour of the paper?

Answer

Since the colour of an object is determined by that frequency of light that is *reflected*, the sheet of paper will appear green, as this is the only frequency that is reflected. All the other frequencies are absorbed by the paper.



Worked Example 61: Colour of an opaque object II

Question: The cover of a book appears to have a magenta colour. What colours of light does it reflect and what colours does it absorb?

Answer

We know that magenta is a combination of red and blue primary colours of light. Therefore the object must be reflecting blue and red light and absorb green.

10.4.2 Colour of transparent objects

If an object is **transparent** it means that you can see through it. For example, glass, clean water and some clear plastics are transparent. The colour of a transparent object is determined by the colours (frequencies) of light which it *transmits* (allows to pass through it). For example, a cup made of green glass will appear green because it absorbs all the other frequencies of light *except* green, which it transmits. This is the light which we receive in our eyes and the object appears green.


Worked Example 62: Colour of Transparent Objects

Question: If white light is shone through a glass plate that absorbs light of all frequencies except red, what is the colour of the glass plate?

Answer

Since the colour of an object is determined by that frequency of light that is *transmitted*, the glass plate will appear red, as this is the only frequency that is not absorbed.

10.4.3 Pigment primary colours

The primary pigments and paints are **cyan**, **magenta** and **yellow**. When pigments or paints of these three colours are mixed together in equal amounts they produce **black**. Any other colour of paint can be made by mixing the primary pigments together in different quantities. The primary pigments are related to the primary colours of light in the following way:

PRIMARY PIGMENTS		
cyan + magenta + yellow = black		
PRIMARY PIGMENTS		PRIMARY COLOURS OF LIGHT
cyan + magenta	=	blue
cyan + yellow	=	green
magenta + yellow	=	red



Colour printers only use 4 colours of ink: cyan, magenta, yellow and black. All the other colours can be mixed from these!


Worked Example 63: Pigments

Question: What colours of light are absorbed by a green pigment?

Answer

If the pigment is green, then green light must be *reflected*. Therefore, red and blue light are absorbed.



Worked Example 64: Primary pigments

Question: I have a ruler which reflects red light and absorbs all other colours of light. What colour does the ruler appear in white light? What primary pigments must have been mixed to make the pigment which gives the ruler its colour?

Answer**Step 1 : What is being asked and what are we given?**

We need to determine the colour of the ruler and the pigments which were mixed to make the colour.

Step 2 : An opaque object appears the colour of the light it reflects

The ruler reflects red light and absorbs all other colours. Therefore the ruler appears to be red.

Step 3 : What pigments need to be mixed to get red?

Red pigment is produced when magenta and yellow pigments are mixed. Therefore magenta and yellow pigments were mixed to make the red pigment which gives the ruler its colour.

**Worked Example 65: Paint Colours**

Question: If cyan light shines on a dress that contains a pigment that is capable of absorbing blue, what colour does the dress appear?

Answer**Step 1 : Determine the component colours of cyan light**

Cyan light is made up of blue and green light.

Step 2 : Determine solution

If the dress absorbs the blue light then the green light must be reflected, so the dress will appear green!

10.5 End of Chapter Exercises

- Calculate the wavelength of light which has a frequency of 570×10^{12} Hz.
- Calculate the frequency of light which has a wavelength of 580 nm.
- Complete the following sentence: When white light is dispersed by a prism, light of the colour ? is refracted the most and light of colour ? is refracted the least.
- What are the two types of photoreceptor found in the retina of the human eye called and which type is sensitive to colours?
- What color do the following shirts appear to the human eye when the lights in a room are turned off and the room is completely dark?
 - red shirt
 - blue shirt
 - green shirt
- Two light bulbs, each of a different colour, shine on a sheet of white paper. Each light bulb can be a primary colour of light - red, green, and blue. Depending on which primary colour of light is used, the paper will appear a different color. What colour will the paper appear if the lights are:
 - red and blue?

- (b) red and green?
- (c) green and blue?

7. Match the primary colour of light on the left to its complementary colour on the right:

Column A	Column B
red	yellow
green	cyan
blue	magenta

8. Which combination of colours of light gives magenta?

- A red and yellow
- B green and red
- C blue and cyan
- D blue and red

9. Which combination of colours of light gives cyan?

- A yellow and red
- B green and blue
- C blue and magenta
- D blue and red

10. If yellow light falls on an object whose pigment absorbs green light, what colour will the object appear?

11. If yellow light falls on a blue pigment, what colour will it appear?

Chapter 11

2D and 3D Wavefronts - Grade 12

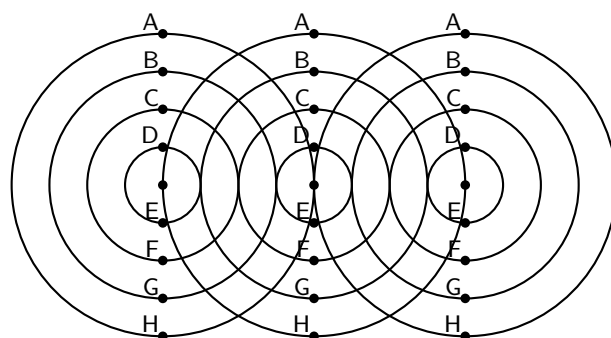
11.1 Introduction

You have learnt about the basic principles of reflection and refraction. In this chapter, you will learn about phenomena that arise with waves in two and three dimensions: interference and diffraction.

11.2 Wavefronts

Activity :: Investigation : Wavefronts

The diagram shows three identical waves being emitted by three point sources. All points marked with the same letter are in phase. Join all points with the same letter.



What type of lines (straight, curved, etc) do you get? How does this compare to the line that joins the sources?

Consider three point sources of waves. If each source emits waves isotropically (i.e. the same in all directions) we will get the situation shown in as shown in Figure 11.1.

We define a **wavefront** as the imaginary line that joins waves that are in phase. These are indicated by the grey, vertical lines in Figure 11.1. The points that are in phase can be peaks, troughs or anything in between, it doesn't matter which points you choose as long as they are in phase.

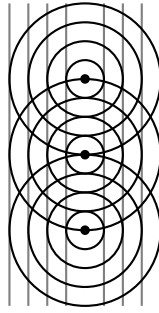


Figure 11.1: Wavefronts are imaginary lines joining waves that are in phase. In the example, the wavefronts (shown by the grey, vertical lines) join all waves at the crest of their cycle.

11.3 The Huygens Principle

Christiaan Huygens described how to determine the path of waves through a medium.



Definition: The Huygens Principle

Each point on a wavefront acts like a point source of circular waves. The waves emitted from these point sources interfere to form another wavefront.

A simple example of the Huygens Principle is to consider the single wavefront in Figure 11.2.

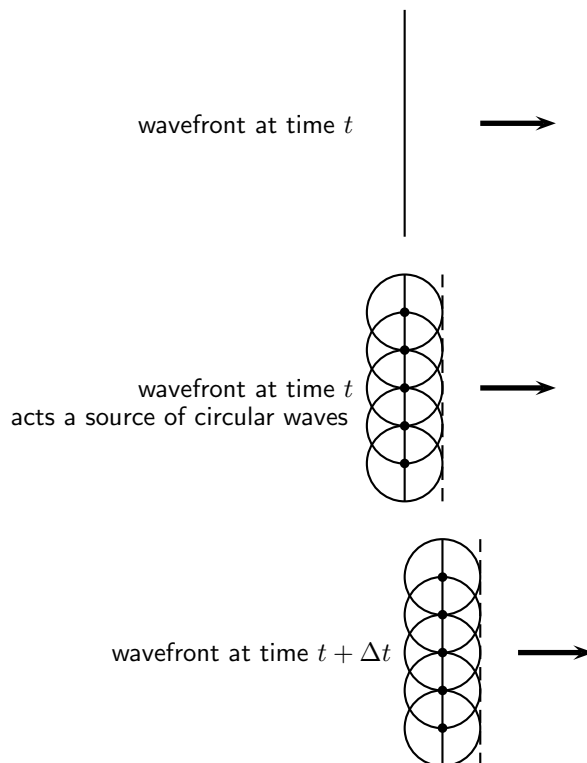
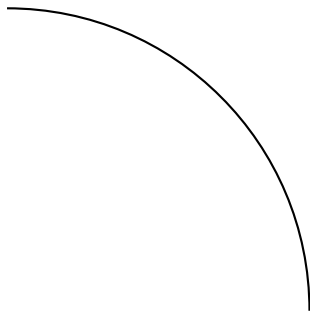


Figure 11.2: A single wavefront at time t acts as a series of point sources of circular waves that interfere to give a new wavefront at a time $t + \Delta t$. The process continues and applies to any shape of waveform.



Worked Example 66: Application of the Huygens Principle

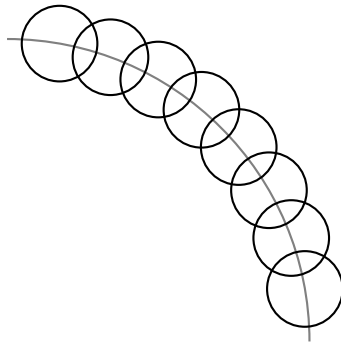
Question: Given the wavefront,



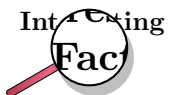
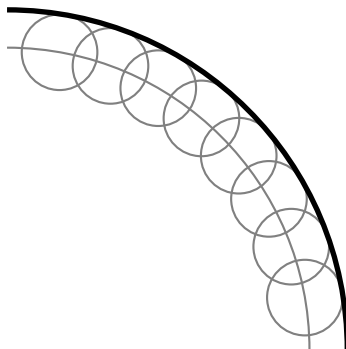
use the Huygens Principle to determine the wavefront at a later time.

Answer

Step 1 : Draw circles at various points along the given wavefront



Step 2 : Join the crests of each circle to get the wavefront at a later time



Christiaan Huygens (14 April 1629 - 8 July 1695), was a Dutch mathematician, astronomer and physicist; born in The Hague as the son of Constantijn Huygens. He studied law at the University of Leiden and the College of Orange in Breda before turning to science. Historians commonly associate Huygens with the scientific revolution.

Huygens generally receives minor credit for his role in the development of modern calculus. He also achieved note for his arguments that light consisted of waves; see: wave-particle duality in Chapter ???. In 1655, he discovered Saturn's moon Titan. He also examined Saturn's planetary rings, and in 1656 he discovered that those rings consisted of rocks. In the same year he observed and sketched the Orion Nebula. He also discovered several interstellar nebulae and some double stars.

11.4 Interference

Interference occurs when two identical waves pass through the same region of space at the same time resulting in a superposition of waves. There are two types of interference which is of interest: **constructive** interference and **destructive** interference.

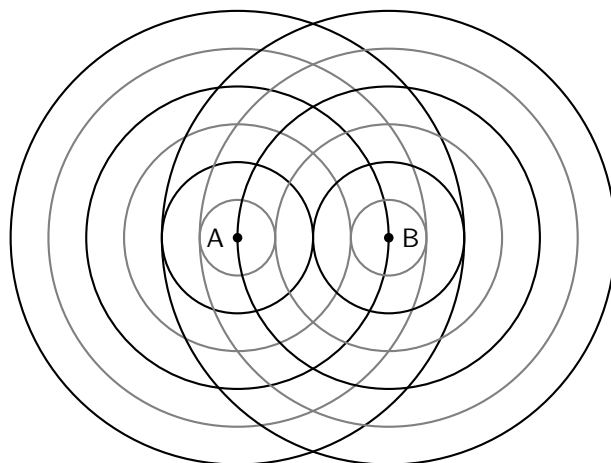
Constructive interference occurs when both waves have a displacement in the same direction, while destructive interference occurs when one wave has a displacement in the opposite direction to the other, thereby resulting in a cancellation. When two waves interfere destructively, the resultant absolute displacement of the medium is less than in either of the individual displacements. When *total* destructive interference occurs, there is no displacement of the medium. For constructive interference the displacement of the medium is greater than the individual displacements.

Constructive interference is the result of two waves with similar phase overlapping. This means that positive parts of one wave tend to overlap with positive parts of the other, and alike for the negative parts. When positive is added to positive and negative is added to negative, the net absolute displacement of the medium is greater than the displacements of the individual waves.

Destructive interference, on the other hand, is the result of two waves with non-similar phase (i.e. anti-phase) overlapping. This means that the positive parts of one wave tend to align with the negative parts of the second wave. When the waves are added together, the positive and negative contributions lead to a net absolute displacement of the medium which is less than the absolute displacements of either of the individual waves. A place where destructive interference takes place is called a node.

Waves can interfere at places where there is never a trough and trough or peak and peak or trough and peak at the same time. At these places the waves will add together and the resultant displacement will be the sum of the two waves but they won't be points of maximum interference.

Consider the two identical waves shown in the picture below. The wavefronts of the peaks are shown as black lines while the wavefronts of the troughs are shown as grey lines. You can see that the black lines cross other black lines in many places. This means two peaks are in the same place at the same time so we will have constructive interference where the two peaks add together to form a bigger peak.



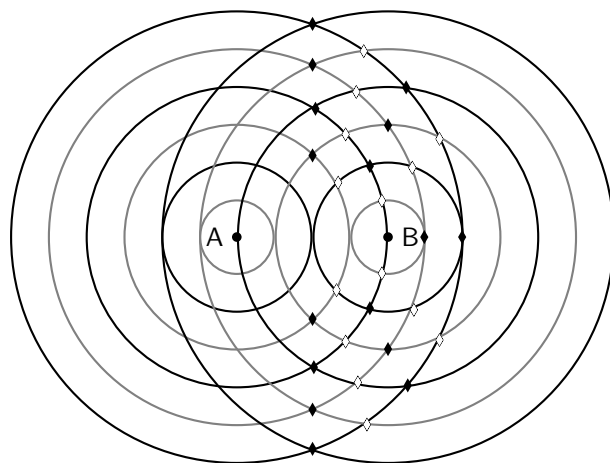
Two points sources (A and B) radiate identical waves. The wavefronts of the peaks (black lines) and troughs (grey lines) are shown. Constructive interference occurs where two black lines intersect or where two grey lines intersect. Destructive interference occurs where a black line intersects with a grey line.

When the grey lines cross other grey lines there are two troughs in the same place at the same time so we will have constructive interference where the two troughs add together to form a bigger trough.

In the case where a grey line crosses a black line we are seeing a trough and peak at the same time. These will cancel each other out and the medium will have no displacement at that point.

- black line + black line = peak + peak = constructive interference
- grey line + grey line = trough + trough = constructive interference
- black line + grey line = grey line + black line = peak + trough = trough + peak = destructive interference

On half the picture below, we have marked the constructive interference with a solid black diamond and the destructive interference with a hollow diamond.



To see if you understand it, cover up the half we have marked with diamonds and try to work out which points are constructive and destructive on the other half of the picture. The two halves are mirror images of each other so you can check yourself.

11.5 Diffraction

One of the most interesting, and also very useful, properties of waves is **diffraction**.



Definition: Diffraction

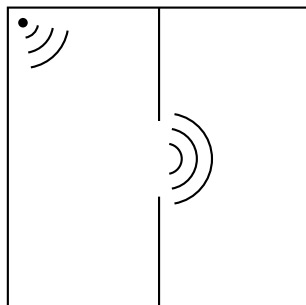
Diffraction is the ability of a wave to spread out in wavefronts as the wave passes through a small aperture or around a sharp edge.



Extension: Diffraction

Diffraction refers to various phenomena associated with wave propagation, such as the bending, spreading and interference of waves emerging from an aperture. It occurs with any type of wave, including sound waves, water waves, electromagnetic waves such as light and radio waves. While diffraction always occurs, its effects are generally only noticeable for waves where the wavelength is on the order of the feature size of the diffracting objects or apertures.

For example, if two rooms are connected by an open doorway and a sound is produced in a remote corner of one of them, a person in the other room will hear the sound as if it originated at the doorway.



As far as the second room is concerned, the vibrating air in the doorway is the source of the sound. The same is true of light passing the edge of an obstacle, but this is not as easily observed because of the short wavelength of visible light.

This means that when waves move through small holes they appear to bend around the sides because there are not enough points on the wavefront to form another straight wavefront. This is bending round the sides we call *diffraction*.



Extension: Diffraction

Diffraction effects are more clear for water waves with longer wavelengths. Diffraction can be demonstrated by placing small barriers and obstacles in a ripple tank and observing the path of the water waves as they encounter the obstacles. The waves are seen to pass around the barrier into the regions behind it; subsequently the water behind the barrier is disturbed. The amount of diffraction (the sharpness of the bending) increases with increasing wavelength and decreases with decreasing wavelength. In fact, when the wavelength of the waves are smaller than the obstacle, no noticeable diffraction occurs.

Activity :: Experiment : Diffraction

Water waves in a ripple tank can be used to demonstrate diffraction and interference.

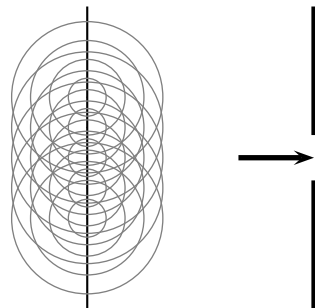
- Turn on the wave generator so that it produces waves with a high frequency (short wavelength).
 - Place a few obstacles, one at a time, (e.g. a brick or a ruler) in the ripple tank. What happens to the wavefronts as they propagate near/past the obstacles? Draw your observations.
 - How does the diffraction change when you change the size of the object?
- Now turn down the frequency of the wave generator so that it produces waves with longer wavelengths.
 - Place the same obstacles in the ripple tank (one at a time). What happens to the wavefronts as they propagate near/past the obstacles? Draw your observations.
 - How does the diffraction change from the higher frequency case?

- Remove all obstacles from the ripple tank and insert a second wave generator. Turn on both generators so that they start at the same time and have the same frequency.
 - What do you notice when the two sets of wavefronts meet each other?
 - Can you identify regions of constructive and destructive interference?
 - Now turn on the generators so that they are out of phase (i.e. start them so that they do not make waves at exactly the same time).
 - What do you notice when the two sets of wavefronts meet each other?
 - Can you identify regions of constructive and destructive interference?
-

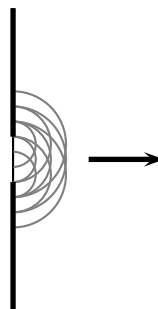
11.5.1 Diffraction through a Slit

When a wave strikes a barrier with a hole only part of the wave can move through the hole. If the hole is similar in size to the wavelength of the wave then diffraction occurs. The waves that come through the hole no longer look like a straight wave front. It bends around the edges of the hole. If the hole is small enough it acts like a point source of circular waves.

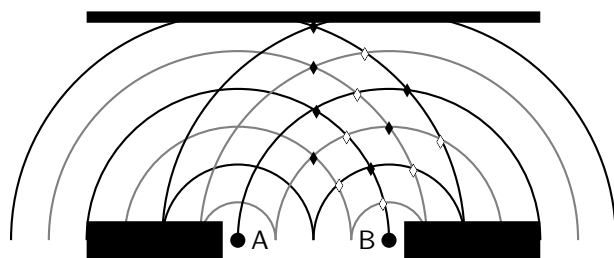
Now if we allow the wavefront to impinge on a barrier with a hole in it, then only the points on the wavefront that move into the hole can continue emitting forward moving waves - but because a lot of the wavefront has been removed, the points on the edges of the hole emit waves that bend round the edges.



If you employ Huygens' principle you can see the effect is that the wavefronts are no longer straight lines.

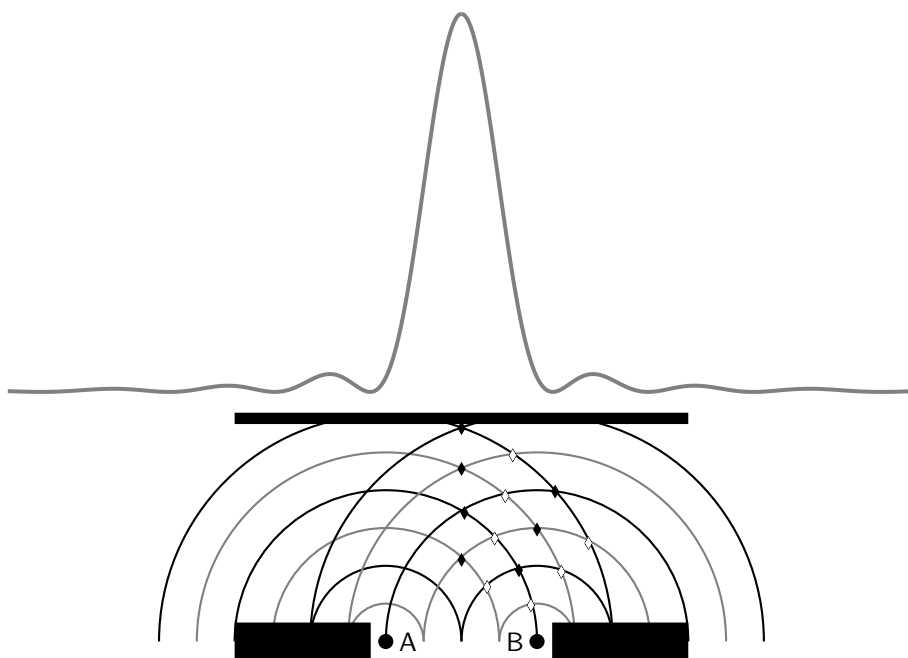


Each point of the slit acts like a point source. If we think about the two point sources on the edges of the slit and call them A and B then we can go back to the diagram we had earlier but with some parts blocked by the wall.



If this diagram were showing sound waves then the sound would be louder (constructive interference) in some places and quieter (destructive interference) in others. You can start to see that there will be a pattern (interference pattern) to the louder and quieter places. If we were studying light waves then the light would be brighter in some places than others depending on the interference.

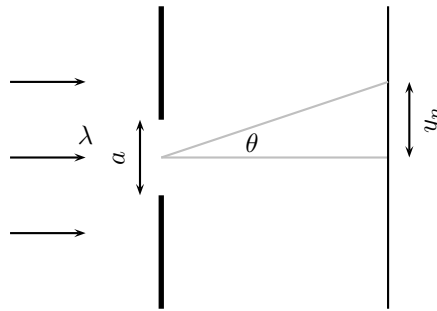
The intensity (how bright or loud) of the interference pattern for a single narrow slit looks like this:



The picture above shows how the waves add together to form the interference pattern. The peaks correspond to places where the waves are adding most intensely and the minima are places where destructive interference is taking place. When looking at interference patterns from light the spectrum looks like:



There is a formula we can use to determine where the peaks and minima are in the interference spectrum. There will be more than one minimum. There are the same number of minima on either side of the central peak and the distances from the first one on each side are the same to the peak. The distances to the peak from the second minimum on each side is also the same, in fact the two sides are mirror images of each other. We label the first minimum that corresponds to a positive angle from the centre as $m = 1$ and the first on the other side (a negative angle from the centre) as $m = -1$, the second set of minima are labelled $m = 2$ and $m = -2$ etc.



The equation for the angle at which the minima occur is given in the definition below:



Definition: Interference Minima

The angle at which the minima in the interference spectrum occur is:

$$\sin \theta = \frac{m\lambda}{a}$$

where

θ is the angle to the minimum

a is the width of the slit

λ is the wavelength of the impinging wavefronts

m is the order of the minimum, $m = \pm 1, \pm 2, \pm 3, \dots$



Worked Example 67: Diffraction Minimum I

Question: A slit with a width of 2511 nm has red light of wavelength 650 nm impinge on it. The diffracted light interferes on a surface. At which angle will the first minimum be?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. The slit has a width of 2511 nm which is 2511×10^{-9} m and we know that the wavelength of the light is 650 nm which is 650×10^{-9} m. We are looking to determine the angle to first minimum so we know that $m = 1$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin \theta = \frac{m\lambda}{a}$$

We can use this relationship to find the angle to the minimum by substituting what we know and solving for the angle.

Step 3 : Substitution

$$\begin{aligned} \sin \theta &= \frac{650 \times 10^{-9} \text{m}}{2511 \times 10^{-9} \text{m}} \\ \sin \theta &= \frac{650}{2511} \\ \sin \theta &= 0.258861012 \\ \theta &= \sin^{-1} 0.258861012 \\ \theta &= 15^\circ \end{aligned}$$

The first minimum is at 15 degrees from the centre peak.


Worked Example 68: Diffraction Minimum II

Question: A slit with a width of 2511 nm has green light of wavelength 532 nm impinge on it. The diffracted light interferes on a surface, at what angle will the first minimum be?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. The slit has a width of 2511 nm which is 2511×10^{-9} m and we know that the wavelength of the light is 532 nm which is 532×10^{-9} m. We are looking to determine the angle to first minimum so we know that $m = 1$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin \theta = \frac{m\lambda}{a}$$

We can use this relationship to find the angle to the minimum by substituting what we know and solving for the angle.

Step 3 : Substitution

$$\begin{aligned} \sin \theta &= \frac{532 \times 10^{-9} \text{m}}{2511 \times 10^{-9} \text{m}} \\ \sin \theta &= \frac{532}{2511} \\ \sin \theta &= 0.211867782 \\ \theta &= \sin^{-1} 0.211867782 \\ \theta &= 12.2^\circ \end{aligned}$$

The first minimum is at 12.2 degrees from the centre peak.

From the formula $\sin \theta = \frac{m\lambda}{a}$ you can see that a smaller wavelength for the same slit results in a smaller angle to the interference minimum. This is something you just saw in the two worked examples. Do a sanity check, go back and see if the answer makes sense. Ask yourself which light had the longer wavelength, which light had the larger angle and what do you expect for longer wavelengths from the formula.


Worked Example 69: Diffraction Minimum III

Question: A slit has a width which is unknown and has green light of wavelength 532 nm impinge on it. The diffracted light interferes on a surface, and the first minimum is measured at an angle of 20.77 degrees?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. We know that the wavelength of the light is 532 nm which is 532×10^{-9} m. We know the angle to first minimum so we know that $m = 1$ and $\theta = 20.77^\circ$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin \theta = \frac{m\lambda}{a}$$

We can use this relationship to find the width by substituting what we know and solving for the width.

Step 3 : Substitution

$$\begin{aligned} \sin \theta &= \frac{532 \times 10^{-9} \text{ m}}{a} \\ \sin 20.77^\circ &= \frac{532 \times 10^{-9}}{a} \\ a &= \frac{532 \times 10^{-9}}{0.354666667} \\ a &= 1500 \times 10^{-9} \\ a &= 1500 \text{ nm} \end{aligned}$$

The slit width is 1500 nm.

11.6 Shock Waves and Sonic Booms

Now we know that the waves move away from the source at the speed of sound. What happens if the source moves at the same time as emitting sounds? Once a sound wave has been emitted it is no longer connected to the source so if the source moves it doesn't change the way the sound wave is propagating through the medium. This means a source can actually catch up to the sound waves it has emitted.

The speed of sound is very fast in air, about $340 \text{ m} \cdot \text{s}^{-1}$, so if we want to talk about a source catching up to sound waves then the source has to be able to move very fast. A good source of sound waves to discuss is a jet aircraft. Fighter jets can move very fast and they are very noisy so they are a good source of sound for our discussion. Here are the speeds for a selection of aircraft that can fly faster than the speed of sound.

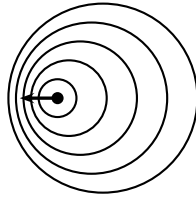
Aircraft	speed at altitude ($\text{km} \cdot \text{h}^{-1}$)	speed at altitude ($\text{m} \cdot \text{s}^{-1}$)
Concorde	2 330	647
Gripen	2 410	669
Mirage F1	2 573	715
Mig 27	1 885	524
F 15	2 660	739
F 16	2 414	671

11.6.1 Subsonic Flight

**Definition: Subsonic**

Subsonic refers to speeds slower than the speed of sound.

When a source emits sound waves and is moving slower than the speed of sound you get the situation in this picture. Notice that the source moving means that the wavefronts, and therefore peaks in the wave, are actually closer together in the one direction and further apart in the other.



subsonic flight

If you measure the waves on the side where the peaks are closer together you'll measure a different wavelength than on the other side of the source. This means that the noise from the source will sound different on the different sides. This is called the *Doppler Effect*.

**Definition: Doppler Effect**

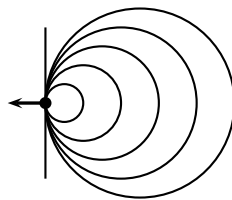
when the wavelength and frequency measured by an observer are different to those emitted by the source due to movement of the source or observer.

11.6.2 Supersonic Flight

**Definition: Supersonic**

Supersonic refers to speeds faster than the speed of sound.

If a plane flies at exactly the speed of sound then the waves that it emits in the direction it is flying won't be able to get away from the plane. It also means that the next sound wave emitted will be exactly on top of the previous one, look at this picture to see what the wavefronts would look like:



shock wave at Mach 1

Sometimes we use the speed of sound as a reference to describe the speed of the object (aircraft in our discussion).

**Definition: Mach Number**

The Mach Number is the ratio of the speed of an object to the speed of sound in the surrounding medium.

Mach number tells you how many times faster than sound the aircraft is moving.

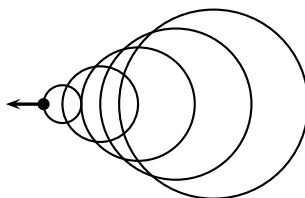
- Mach Number < 1 : aircraft moving slower than the speed of sound
- Mach Number $= 1$: aircraft moving at the speed of sound
- Mach Number > 1 : aircraft moving faster than the speed of sound

To work out the Mach Number divide the speed of the aircraft by the speed of sound.

$$\text{Mach Number} = \frac{v_{\text{aircraft}}}{v_{\text{sound}}}$$

Remember: the units must be the same before you divide.

If the aircraft is moving faster than the speed of sound then the wavefronts look like this:



supersonic shock wave

If the source moves faster than the speed of sound, a cone of wave fronts is created. This is called a Mach cone. From constructive interference, we know that two peaks that add together form a larger peak. In a Mach cone many, many peaks add together to form a very large peak. This is a sound wave so the large peak is a very, very loud sound wave. This sounds like a huge "boom" and we call the noise a *sonic boom*.



Definition: Sonic Boom

A sonic boom is the sound heard by an observer as a shockwave passes.



Worked Example 70: Mach Speed I

Question: An aircraft flies at $1300 \text{ km} \cdot \text{h}^{-1}$ and the speed of sound in air is $340 \text{ m} \cdot \text{s}^{-1}$. What is the Mach Number of the aircraft?

Answer

Step 1 : Check what you are given

We know we are dealing with Mach Number. We are given the speed of sound in air, $340 \text{ m} \cdot \text{s}^{-1}$, and the speed of the aircraft, $1300 \text{ km} \cdot \text{h}^{-1}$. The speed of the aircraft is in different units to the speed of sound so we need to convert the units:

$$\begin{aligned} 1300 \text{ km} \cdot \text{h}^{-1} &= 1300 \text{ km} \cdot \text{h}^{-1} \\ 1300 \text{ km} \cdot \text{h}^{-1} &= 1300 \times \frac{1000 \text{ m}}{3600 \text{ s}} \\ 1300 \text{ km} \cdot \text{h}^{-1} &= 361.1 \text{ m} \cdot \text{s}^{-1} \end{aligned}$$

Step 2 : Applicable principles

We know that there is a relationship between the Mach Number, the speed of sound and the speed of the aircraft:

$$\text{Mach Number} = \frac{v_{\text{aircraft}}}{v_{\text{sound}}}$$

We can use this relationship to find the Mach Number.

Step 3 : Substitution

$$\begin{aligned}\text{Mach Number} &= \frac{v_{\text{aircraft}}}{v_{\text{sound}}} \\ \text{Mach Number} &= \frac{361.1}{340} \\ \text{Mach Number} &= 1.06\end{aligned}$$

The Mach Number is 1.06.



Exercise: Mach Number

In this exercise we will determine the Mach Number for the different aircraft in the previous table. To help you get started we have calculated the Mach Number for the Concorde with a speed of sound $v_{\text{sound}} = 340 \text{ ms}^{-1}$.

For the Concorde we know the speed and we know that:

$$\text{Mach Number} = \frac{v_{\text{aircraft}}}{v_{\text{sound}}}$$

For the Concorde this means that

$$\begin{aligned}\text{Mach Number} &= \frac{647}{340} \\ &= 1.9\end{aligned}$$

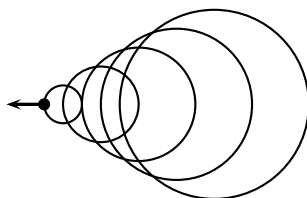
Aircraft	speed at altitude ($\text{km} \cdot \text{h}^{-1}$)	speed at altitude ($\text{m} \cdot \text{s}^{-1}$)	Mach Number
Concorde	2 330	647	1.9
Gripen	2 410	669	
Mirage F1	2 573	715	
Mig 27	1 885	524	
F 15	2 660	739	
F 16	2 414	671	

Now calculate the Mach Numbers for the other aircraft in the table.

11.6.3 Mach Cone

The shape of the Mach Cone depends on the speed of the aircraft. When the Mach Number is 1 there is no cone but as the aircraft goes faster and faster the angle of the cone gets smaller and smaller.

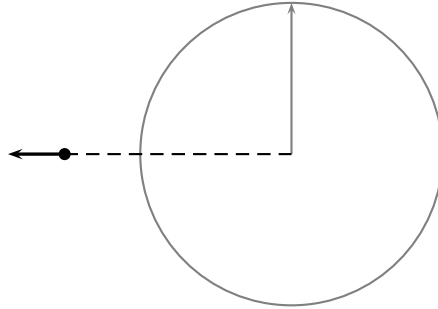
If we go back to the supersonic picture we can work out what the angle of the cone must be.



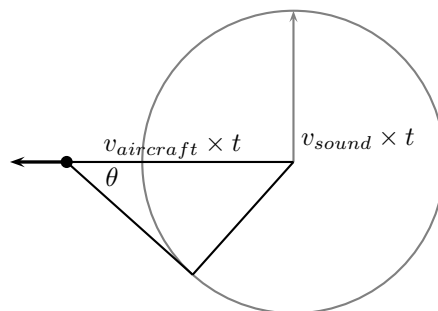
supersonic shock wave

We build a triangle between how far the plane has moved and how far a wavefront at right angles to the direction the plane is flying has moved:

An aircraft emits a sound wavefront. The wavefront moves at the speed of sound $340 \text{ m} \cdot \text{s}^{-1}$ and the aircraft moves at Mach 1.5, which is $1.5 \times 340 = 510 \text{ m} \cdot \text{s}^{-1}$. The aircraft travels faster than the wavefront. If we let the wavefront travel for a time t then the following diagram will apply:



We know how fast the wavefront and the aircraft are moving so we know the distances that they have traveled:



The angle between the cone that forms and the direction of the plane can be found from the right-angle triangle we have drawn into the figure. We know that $\sin \theta = \frac{\text{opposite}}{\text{hypotenuse}}$ which in this figure means:

$$\begin{aligned}\sin \theta &= \frac{\text{opposite}}{\text{hypotenuse}} \\ \sin \theta &= \frac{v_{\text{sound}} \times t}{v_{\text{aircraft}} \times t} \\ \sin \theta &= \frac{v_{\text{sound}}}{v_{\text{aircraft}}}\end{aligned}$$

In this case we have used sound and aircraft but a more general way of saying this is:

- aircraft = source
- sound = wavefront

We often just write the equation as:

$$\begin{aligned}\sin \theta &= \frac{v_{\text{sound}}}{v_{\text{aircraft}}} \\ v_{\text{aircraft}} \sin \theta &= v_{\text{sound}} \\ v_{\text{source}} \sin \theta &= v_{\text{wavefront}} \\ v_s \sin \theta &= v_w\end{aligned}$$

From this equation, we can see that the faster the source (aircraft) moves, the smaller the angle of the Mach cone.



Exercise: Mach Cone

In this exercise we will determine the Mach Cone Angle for the different aircraft in the table mentioned above. To help you get started we have calculated the Mach Cone Angle for the Concorde with a speed of sound $v_{\text{sound}} = 340 \text{ m} \cdot \text{s}^{-1}$.

For the Concorde we know the speed and we know that:

$$\sin \theta = \frac{v_{\text{sound}}}{v_{\text{aircraft}}}$$

For the Concorde this means that

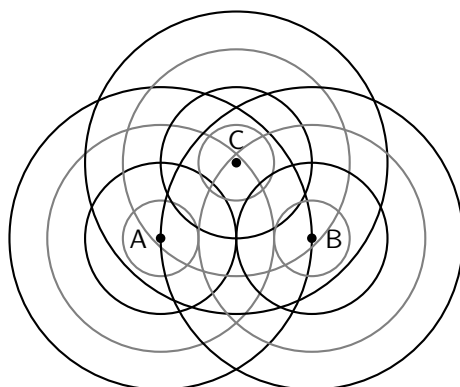
$$\begin{aligned}\sin \theta &= \frac{340}{647} \\ \theta &= \sin^{-1} \frac{340}{647} \\ \theta &= 31.7^\circ\end{aligned}$$

Aircraft	speed at altitude ($\text{km} \cdot \text{h}^{-1}$)	speed at altitude ($\text{m} \cdot \text{s}^{-1}$)	Mach Cone Angle (degrees)
Concorde	2 330	647	31.7
Gripen	2 410	669	
Mirage F1	2 573	715	
Mig 27	1 885	524	
F 15	2 660	739	
F 16	2 414	671	

Now calculate the Mach Cone Angles for the other aircraft in the table.

11.7 End of Chapter Exercises

- In the diagram below the peaks of wavefronts are shown by black lines and the troughs by grey lines. Mark all the points where constructive interference between two waves is taking place and where destructive interference is taking place. Also note whether the interference results in a peak or a trough.



2. For a slit of width 1300 nm, calculate the first 3 minima for light of the following wavelengths:
 - (a) blue at 475 nm
 - (b) green at 510 nm
 - (c) yellow at 570 nm
 - (d) red at 650 nm
3. For light of wavelength 540 nm, determine what the width of the slit needs to be to have the first minimum at:
 - (a) 7.76 degrees
 - (b) 12.47 degrees
 - (c) 21.1 degrees
4. For light of wavelength 635 nm, determine what the width of the slit needs to be to have the second minimum at:
 - (a) 12.22 degrees
 - (b) 18.51 degrees
 - (c) 30.53 degrees
5. If the first minimum is at 8.21 degrees and the second minimum is at 16.6 degrees, what is the wavelength of light and the width of the slit? (**Hint:** solve simultaneously.)
6. Determine the Mach Number, with a speed of sound of $340 \text{ m} \cdot \text{s}^{-1}$, for the following aircraft speeds:
 - (a) $640 \text{ m} \cdot \text{s}^{-1}$
 - (b) $980 \text{ m} \cdot \text{s}^{-1}$
 - (c) $500 \text{ m} \cdot \text{s}^{-1}$
 - (d) $450 \text{ m} \cdot \text{s}^{-1}$
 - (e) $1300 \text{ km} \cdot \text{h}^{-1}$
 - (f) $1450 \text{ km} \cdot \text{h}^{-1}$
 - (g) $1760 \text{ km} \cdot \text{h}^{-1}$
7. If an aircraft has a Mach Number of 3.3 and the speed of sound is $340 \text{ m} \cdot \text{s}^{-1}$, what is its speed?
8. Determine the Mach Cone angle, with a speed of sound of $340 \text{ m} \cdot \text{s}^{-1}$, for the following aircraft speeds:

- (a) $640 \text{ m} \cdot \text{s}^{-1}$
 - (b) $980 \text{ m} \cdot \text{s}^{-1}$
 - (c) $500 \text{ m} \cdot \text{s}^{-1}$
 - (d) $450 \text{ m} \cdot \text{s}^{-1}$
 - (e) $1300 \text{ km} \cdot \text{h}^{-1}$
 - (f) $1450 \text{ km} \cdot \text{h}^{-1}$
 - (g) $1760 \text{ km} \cdot \text{h}^{-1}$
9. Determine the aircraft speed, with a speed of sound of $340 \text{ m} \cdot \text{s}^{-1}$, for the following Mach Cone Angles:
- (a) 58.21 degrees
 - (b) 49.07 degrees
 - (c) 45.1 degrees
 - (d) 39.46 degrees
 - (e) 31.54 degrees

Chapter 12

Wave Nature of Matter - Grade 12

12.1 Introduction

In chapters 15 and 16 the so-called wave-particle duality of light is described. This duality states that light displays properties of both waves and of particles, depending on the experiment performed. For example, interference and diffraction of light are properties of its wave nature, while the photoelectric effect is a property of its particle nature. In fact we call a particle of light a photon.

Hopefully you have realised that nature loves symmetry. So, if light which was originally believed to be a wave also has a particle nature, then perhaps particles, also display a wave nature. In other words matter which we originally thought of as particles may also display a wave-particle duality.

12.2 de Broglie Wavelength

Einstein showed that for a photon, its momentum, p , is equal to its energy, E divided by the speed of light, c :

$$p = \frac{E}{c}.$$

The energy of the photon can also be expressed in terms of the wavelength of the light, λ :

$$E = \frac{hc}{\lambda},$$

where h is Planck's constant. Combining these two equations we find that the the momentum of the photon is related to its wavelength

$$p = \frac{hc}{c\lambda} = \frac{h}{\lambda},$$

or equivalently

$$\lambda = \frac{h}{p}.$$

In 1923, Louis de Broglie proposed that this equation not only holds for photons, but also holds for particles of matter. This is known as the de Broglie hypothesis.

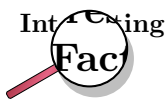
**Definition: De Broglie Hypothesis**

A particle of mass m moving with velocity v has a wavelength λ related to its momentum $p = mv$ by

$$\lambda = \frac{h}{p} = \frac{h}{mv} \quad (12.1)$$

This wavelength, λ , is known as the de Broglie wavelength of the particle (where h is Planck's constant).

Since the value of Planck's constant is incredibly small $h = 6,63 \times 10^{-34} \text{ J} \cdot \text{s}$, the wavelike nature of everyday objects is not really observable.



The de Broglie hypothesis was proposed by French physicist Louis de Broglie (15 August 1892 – 19 March 1987) in 1923 in his PhD thesis. He was awarded the Nobel Prize for Physics in 1929 for this work, which made him the first person to receive a Nobel Prize on a PhD thesis.



Worked Example 71: de Broglie Wavelength of a Cricket Ball

Question: A cricket ball has a mass of 0,150 kg and is bowled towards a bowler at $40 \text{ m} \cdot \text{s}^{-1}$. Calculate the de Broglie wavelength of the cricket ball?

Answer

Step 1 : Determine what is required and how to approach the problem

We are required to calculate the de Broglie wavelength of a cricket ball given its mass and speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The mass of the cricket ball $m = 0,150 \text{ kg}$
- The velocity of the cricket ball $v = 40 \text{ m} \cdot \text{s}^{-1}$

and we know:

- Planck's constant $h = 6,63 \times 10^{-34} \text{ J} \cdot \text{s}$

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned} \lambda &= \frac{h}{mv} \\ &= \frac{6,63 \times 10^{-34} \text{ J} \cdot \text{s}}{(0,150 \text{ kg})(40 \text{ m} \cdot \text{s}^{-1})} \\ &= 1,11 \times 10^{-34} \text{ m} \end{aligned}$$

This wavelength is considerably smaller than the diameter of a proton which is approximately 10^{-15} m . Hence the wave-like properties of this cricket ball are too small to be observed.



Worked Example 72: The de Broglie wavelength of an electron

Question: Calculate the de Broglie wavelength of an electron moving at $40 \text{ m} \cdot \text{s}^{-1}$.

Answer**Step 1 : Determine what is required and how to approach the problem**

We are required to calculate the de Broglie wavelength of an electron given its speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The velocity of the electron $v = 40 \text{ m} \cdot \text{s}^{-1}$

and we know:

- The mass of the electron $m_e = 9,11 \times 10^{-31} \text{ kg}$
- Planck's constant $h = 6,63 \times 10^{-34} \text{ J} \cdot \text{s}$

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned} \lambda &= \frac{h}{mv} \\ &= \frac{6,63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9,11 \times 10^{-31} \text{ kg})(40 \text{ m} \cdot \text{s}^{-1})} \\ &= 1,82 \times 10^{-5} \text{ m} \\ &= 0,0182 \text{ mm} \end{aligned}$$

Although the electron and cricket ball in the two previous examples are travelling at the same velocity the de Broglie wavelength of the electron is much larger than that of the cricket ball. This is because the wavelength is inversely proportional to the mass of the particle.

**Worked Example 73: The de Broglie wavelength of an electron**

Question: Calculate the de Broglie wavelength of a electron moving at $3 \times 10^5 \text{ m} \cdot \text{s}^{-1}$. ($\frac{1}{1000}$ of the speed of light.)

Answer**Step 1 : Determine what is required and how to approach the problem**

We are required to calculate the de Broglie wavelength of an electron given its speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The velocity of the electron $v = 3 \times 10^5 \text{ m} \cdot \text{s}^{-1}$

and we know:

- The mass of the electron $m = 9,11 \times 10^{-31} \text{ kg}$
- Planck's constant $h = 6,63 \times 10^{-34} \text{ J} \cdot \text{s}$

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned}
 \lambda &= \frac{h}{mv} \\
 &= \frac{6,63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9,11 \times 10^{-31} \text{ kg})(3 \times 10^5 \text{ m} \cdot \text{s}^{-1})} \\
 &= 2,43 \times 10^{-9} \text{ m}
 \end{aligned}$$

This is the size of an atom. For this reason, electrons moving at high velocities can be used to “probe” the structure of atoms. This is discussed in more detail at the end of this chapter. Figure 12.1 compares the wavelengths of fast moving electrons to the wavelengths of visible light.

Since the de Broglie wavelength of a particle is inversely proportional to its velocity, the wavelength decreases as the velocity increases. This is confirmed in the last two examples with the electrons. De Broglie’s hypothesis was confirmed by Davisson and Germer in 1927 when they observed a beam of electrons being diffracted off a nickel surface. The diffraction means that the moving electrons have a wave nature. They were also able to determine the wavelength of the electrons from the diffraction. To measure a wavelength one needs two or more diffracting centres such as pinholes, slits or atoms. For diffraction to occur the centres must be separated by a distance about the same size as the wavelength. Theoretically, all objects, not just sub-atomic particles, exhibit wave properties according to the de Broglie hypothesis.

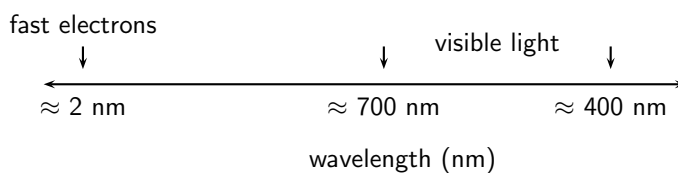


Figure 12.1: The wavelengths of the fast electrons are much smaller than that of visible light.

12.3 The Electron Microscope

We have seen that under certain circumstances particles behave like waves. This idea is used in the electron microscope which is a type of microscope that uses electrons to create an image of the target. It has much higher magnification or resolving power than a normal light microscope, up to two million times, allowing it to see smaller objects and details.

Let’s first review how a regular optical microscope works. A beam of light is shone through a thin target and the image is then magnified and focused using objective and ocular lenses. The amount of light which passes through the target depends on the densities of the target since the less dense regions allow more light to pass through than the denser regions. This means that the beam of light which is partially transmitted through the target carries information about the inner structure of the target.

The original form of the electron microscopy, transmission electron microscopy, works in a similar manner using electrons. In the electron microscope, electrons which are emitted by a cathode are formed into a beam using magnetic lenses. This electron beam is then passed through a very thin target. Again, the regions in the target with higher densities stop the electrons more easily. So, the number of electrons which pass through the different regions of the target depends on their densities. This means that the partially transmitted beam of electrons carries information about the densities of the inner structure of the target. The spatial variation in this information (the “image”) is then magnified by a series of magnetic lenses and it is recorded by hitting a fluorescent screen, photographic plate, or light sensitive sensor such as a CCD (charge-coupled device) camera. The image detected by the CCD may be displayed in real time on a monitor

or computer. In figure 12.2 is an image of the polio virus obtained with a transmission electron microscope.

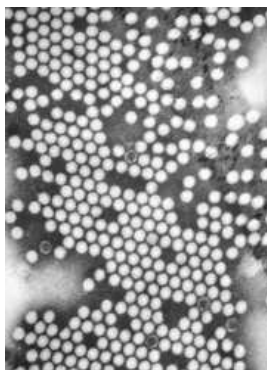


Figure 12.2: The image of the polio virus using a transmission electron microscope.

The structure of an optical and electron microscope are compared in figure 12.3. While the optical microscope uses light and focuses using lenses, the electron microscope uses electrons and focuses using electromagnets.

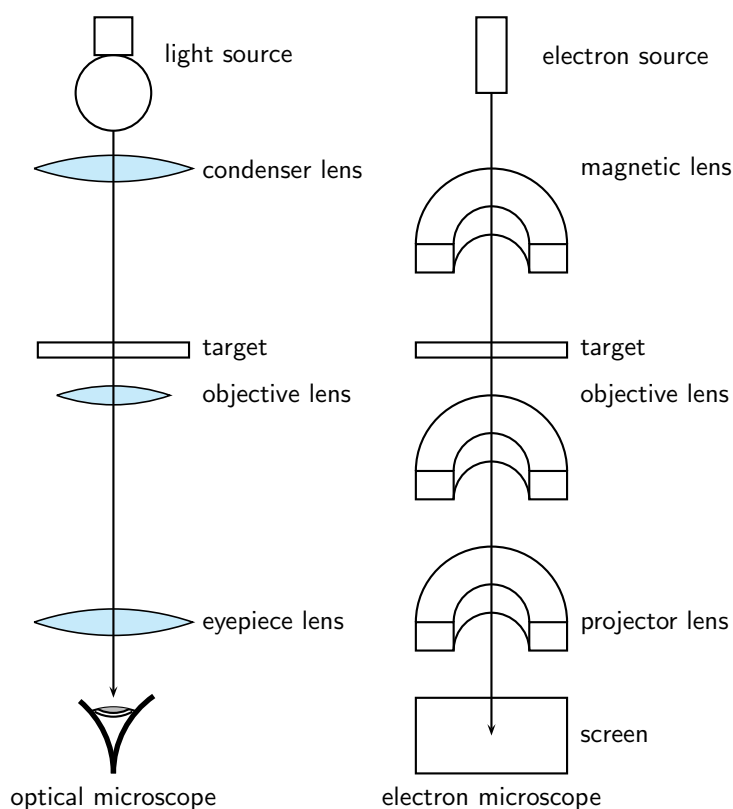


Figure 12.3: Diagram of the basic components of an optical microscope and an electron microscope.

Electron microscopes are very useful as they are able to magnify objects to a much higher resolution. This is because their de Broglie wavelengths are so much smaller than that of visible light. You hopefully remember that light is diffracted by objects which are separated by a distance of about the same size as the wavelength of the light. This diffraction then prevents you from being able to focus the transmitted light into an image. So the sizes at which diffraction occurs for a beam of electrons is much smaller than those for visible light. This is why you can magnify targets to a much higher order of magnification using electrons rather than visible light.

Table 12.1: Comparison of Light and Electron Microscopes

	Light microscope	Electron microscope
Source	Bright lamp or laser	Electron gun
Radiation	U.V. or visible light	Electron beam produced by heating metal surface (e.g. tungsten)
Lenses	Curved glass surfaces	Electromagnets
Receiver	Eye; photographic emulsion or digital image	Fluorescent screen (for location and focusing image); photographic emulsion or digital image
Focus	Axial movement of lenses (up and down)	Adjustment of magnetic field in the electromagnets by changing the current
Operating Pressure	Atmospheric	High vacuum



Extension: High-Resolution Transmission Electron Microscope (HRTEM)

There are high-resolution TEM (HRTEM) which have been built. In fact the resolution is sufficient to show carbon atoms in diamond separated by only 89 picometers and atoms in silicon at 78 picometers. This is at magnifications of 50 million times. The ability to determine the positions of atoms within materials has made the HRTEM a very useful tool for nano-technologies research. It is also very important for the development of semiconductor devices for electronics and photonics.

Transmission electron microscopes produce two-dimensional images.



Extension: Scanning Electron Microscope (SEM)

The Scanning Electron Microscope (SEM) produces images by hitting the target with a primary electron beam which then excites the surface of the target. This causes secondary electrons to be emitted from the surface which are then detected. So the electron beam in the SEM is moved (or scanned) across the sample, while detectors build an image from the secondary electrons.

Generally, the transmission electron microscope's resolution is about an order of magnitude better than the SEM resolution. However, because the SEM image relies on surface processes rather than transmission it is able to image bulk samples (unlike optical microscopes and TEM which require the samples to be thin) and has a much greater depth of view, and so can produce images that are a good representation of the 3D structure of the sample.

12.3.1 Disadvantages of an Electron Microscope

Electron microscopes are expensive to buy and maintain. They are also very sensitive to vibration and external magnetic fields. This means that special facilities are required to house microscopes aimed at achieving high resolutions. Also the targets have to be viewed in vacuum, as the electrons would scatter off the molecules that make up air.

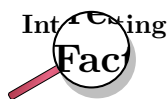


Extension: Scanning Electron Microscope (SEM)

Scanning electron microscopes usually image conductive or semi-conductive materials best. A common preparation technique is to coat the target with a several-nanometer layer of conductive material, such as gold, from a sputtering machine; however this process has the potential to disturb delicate samples.

The targets have to be prepared in many ways to give proper detail. This may result in artifacts purely as a result of the treatment. This gives the problem of

distinguishing artifacts from material, particularly in biological samples. Scientists maintain that the results from various preparation techniques have been compared, and as there is no reason that they should all produce similar artifacts, it is therefore reasonable to believe that electron microscopy features correlate with living cells.



The first electron microscope prototype was built in 1931 by the German engineers Ernst Ruska and Max Knoll. It was based on the ideas and discoveries of Louis de Broglie. Although it was primitive and was not ideal for practical use, the instrument was still capable of magnifying objects by four hundred times. The first practical electron microscope was built at the University of Toronto in 1938, by Eli Franklin Burton and students Cecil Hall, James Hillier and Albert Prebus.

Although modern electron microscopes can magnify objects up to two million times, they are still based upon Ruska's prototype and his correlation between wavelength and resolution. The electron microscope is an integral part of many laboratories. Researchers use it to examine biological materials (such as microorganisms and cells), a variety of large molecules, medical biopsy samples, metals and crystalline structures, and the characteristics of various surfaces.

12.3.2 Uses of Electron Microscopes

Electron microscopes can be used to study:

- the topography of an object – how its surface looks.
- the morphology of particles making up an object – their shapes and sizes.
- the composition of an object – the elements and compounds that the object is composed of and the relative amounts of them.
- the crystallographic information for crystalline samples – how the atoms are arranged in the object.

12.4 End of Chapter Exercises

1. If the following particles have the same velocity, which has the shortest wavelength: electron, hydrogen atom, lead atom?
2. A bullet weighing 30 g is fired at a velocity of $500 \text{ m} \cdot \text{s}^{-1}$. What is its wavelength?
3. Calculate the wavelength of an electron which has a kinetic energy of $1.602 \times 10^{-19} \text{ J}$.
4. If the wavelength of an electron is 10^{-9} m what is its velocity?
5. Considering how one calculates wavelength using slits, try to explain why we would not be able to physically observe diffraction of the cricket ball in the first worked example.

Chapter 13

Electrodynamics - Grade 12

13.1 Introduction

In Grade 11 you learnt how a magnetic field is generated around a current carrying conductor. You also learnt how a current is generated in a conductor that moves in a magnetic field. This chapter describes how conductors moving in a magnetic field are applied in the real-world.

13.2 Electrical machines - generators and motors

We have seen that when a conductor is moved in a magnetic field or when a magnet is moved near a conductor, a current flows in the conductor. The amount of current depends on the speed at which the conductor experiences a changing magnetic field, the number of turns of the conductor and the position of the plane of the conductor with respect to the magnetic field. The effect of the orientation of the conductor with respect to the magnetic field is illustrated in Figure 13.1.

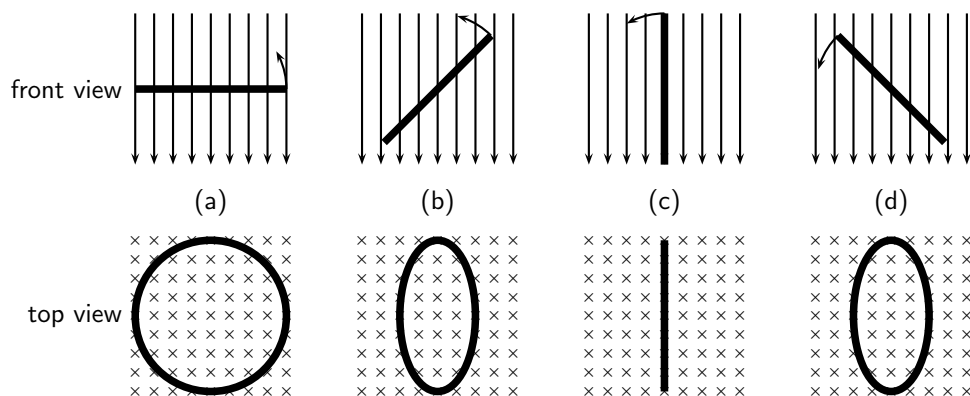


Figure 13.1: Series of figures showing that the magnetic flux through a conductor is dependent on the angle that the plane of the conductor makes with the magnetic field. The greatest flux passes through the conductor when the plane of the conductor is perpendicular to the magnetic field lines as in (a). The number of field lines passing through the conductor decreases, as the conductor rotates until it is parallel to the magnetic field (c).

If the current flowing in the conductor were plotted as a function of the angle between the plane of the conductor and the magnetic field, then the current would vary as shown in Figure 13.2. The current alternates about zero and is known as an *alternating current* (abbreviated AC).

Recall Faraday's Law which you learnt about in Grade 11:

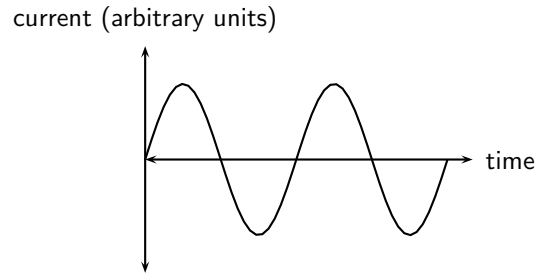


Figure 13.2: Variation of current as the angle of the plane of a conductor with the magnetic field changes.



Definition: Faraday's Law

The emf (electromotive force), ϵ , produced around a loop of conductor is proportional to the rate of change of the magnetic flux, ϕ , through the area, A , of the loop. This can be stated mathematically as:

$$\epsilon = -N \frac{\Delta\phi}{\Delta t}$$

where $\phi = B \cdot A$ and B is the strength of the magnetic field.

Faraday's Law relates induced emf (electromotive force) to the rate of change of flux, which is the product of the magnetic field and the cross-sectional area the field lines pass through. As the closed loop conductor changes orientation with respect to the magnetic field, the amount of magnetic flux through the area of the loop changes, and an emf is induced in the conducting loop.

13.2.1 Electrical generators

AC generator

The principle of rotating a conductor in a magnetic field is used in electrical generators. A generator converts mechanical energy (motion) into electrical energy.



Definition: Generator

A generator converts mechanical energy into electrical energy.

The layout of a simple AC generator is shown in Figure 13.3. The conductor in the shape of a coil is connected to a slip ring. The conductor is then manually rotated in the magnetic field generating an alternating emf. The slip rings are connected to the load via brushes.

If a machine is constructed to rotate a magnetic field around a set of stationary wire coils with the turning of a shaft, AC voltage will be produced across the wire coils as that shaft is rotated, in accordance with Faraday's Law of electromagnetic induction. This is the basic operating principle of an AC generator.

In an AC generator the two ends of the coil are each attached to a slip ring that makes contact with brushes as the coil turns. The direction of the current changes with every half turn of the coil. As one side of the loop moves to the other pole of the magnetic field, the current in the loop changes direction. The two slip rings of the AC generator allow the coil to turn without breaking the connections to the load circuit. This type of current which changes direction is known as alternating current.

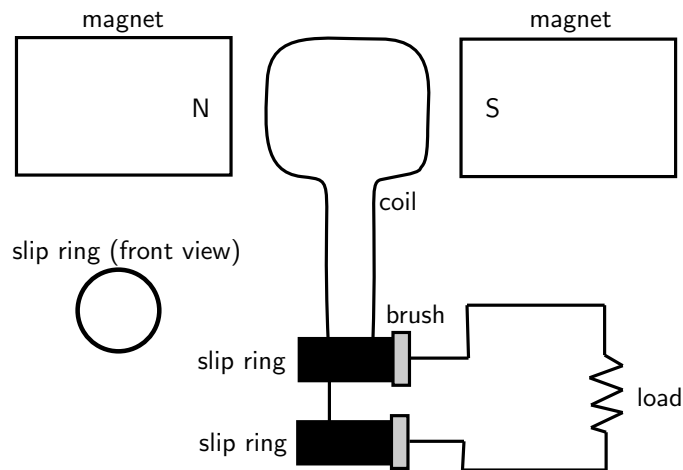


Figure 13.3: Layout of an alternating current generator.



AC generators are also known as alternators. They are found in motor cars to charge the car battery.

DC generator

A simple DC generator is constructed the same way as an AC generator except that there is one slip ring which is split into two pieces, called a commutator, so the current in the external circuit does not change direction. The layout of a DC generator is shown in Figure 13.4. The split-ring commutator accommodates for the change in direction of the current in the loop, thus creating direct current (DC) current going through the brushes and out to the circuit.

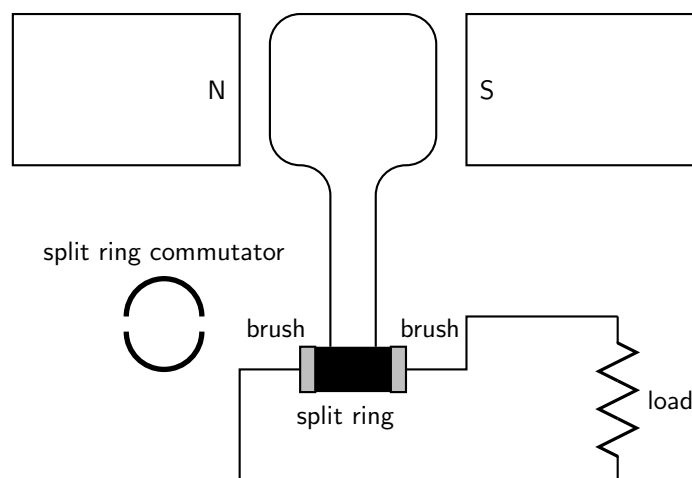


Figure 13.4: Layout of a direct current generator.

The shape of the emf from a DC generator is shown in Figure 13.5. The emf is not steady but is the absolute value of a sine/cosine wave.

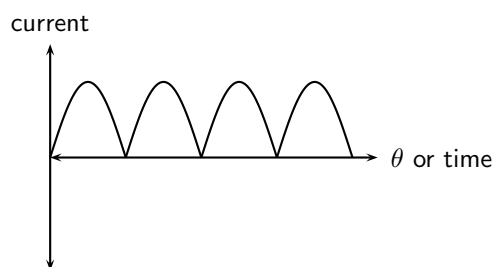


Figure 13.5: Variation of emf in a DC generator.

AC versus DC generators

The problems involved with making and breaking electrical contact with a moving coil are sparking and heat, especially if the generator is turning at high speed. If the atmosphere surrounding the machine contains flammable or explosive vapors, the practical problems of spark-producing brush contacts are even greater.

If the magnetic field, rather than the coil/conductor is rotated, then brushes are not needed in an AC generator (alternator), so an alternator will not have the same problems as DC generators. The same benefits of AC over DC for generator design also apply to electric motors. While DC motors need brushes to make electrical contact with moving coils of wire, AC motors do not. In fact, AC and DC motor designs are very similar to their generator counterparts. The AC motor depends on the reversing magnetic field produced by alternating current through its stationary coils of wire to make the magnet rotate. The DC motor depends on the brush contacts making and breaking connections to reverse current through the rotating coil every 1/2 rotation (180 degrees).

13.2.2 Electric motors

The basic principles of operation for a motor are the same as that of a generator, except that a motor converts electrical energy into mechanical energy (motion).



Definition: Motor

An electric motor converts electrical energy into mechanical energy.

If one were to place a moving charged particle in a magnetic field, it would feel a force called the **Lorentz force**.



Definition: The Lorentz Force

The Lorentz force is the force experienced by a moving charged particle in a magnetic field and can be described by:

$$F = q \times v \times B$$

where

F is the force (in newtons, N)

q is the electric charge (in coulombs, C)

v is the velocity of the charged particle (in $\text{m}\cdot\text{s}^{-1}$) and

B is the magnetic field strength (in teslas, T).

Current in a conductor consists of moving charges. Therefore, a current carrying coil in a magnetic field will also feel the Lorentz force. For a straight current carrying wire which is not moving:

$$F = I \times L \times B$$

where

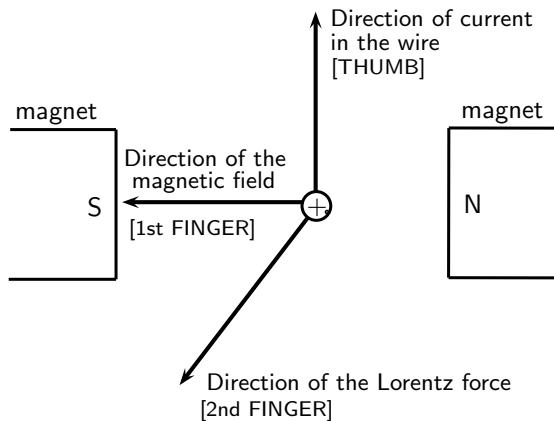
F is the force (in newtons, N)

I is the current in the wire (in amperes, A)

L is the length of the wire which is in the magnetic field (in m) and

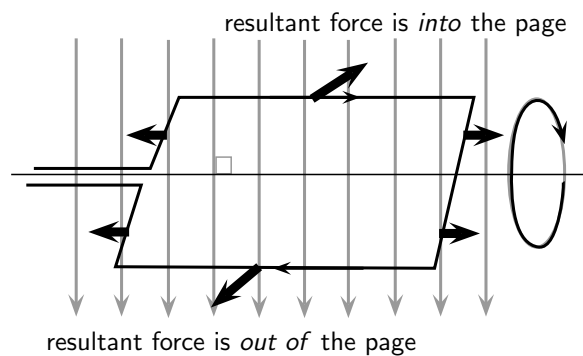
B is the magnetic field strength (in teslas, T).

The direction of the Lorentz force is perpendicular to both the direction of the flow of current and the magnetic field and can be found using the **Right Hand Rule** as shown in the picture below. Use your *right hand*; your thumb points in the direction of the current, your first finger in the direction of the magnetic field and your third finger will then point in the direction of the force.



Both motors and generators can be explained in terms of a coil that rotates in a magnetic field. In a generator the coil is attached to an external circuit that is turned, resulting in a changing flux that induces an emf. In a motor, a current-carrying coil in a magnetic field experiences a force on both sides of the coil, creating a twisting force (called a *torque*, pronounce like 'talk') which makes it turn.

Any coil carrying current can feel a force in a magnetic field. The force is the Lorentz force on the moving charges in the conductor. The force on opposite sides of the coil will be in opposite directions because the charges are moving in opposite directions. This means the coil will rotate, see the picture below:



Instead of rotating the loops through a magnetic field to create electricity, a current is sent through the wires, creating electromagnets. The outer magnets will then repel the electromagnets and rotate the shaft as an electric motor. If the current is AC, the two slip rings are required to create an AC motor. An AC motor is shown in Figure 13.6

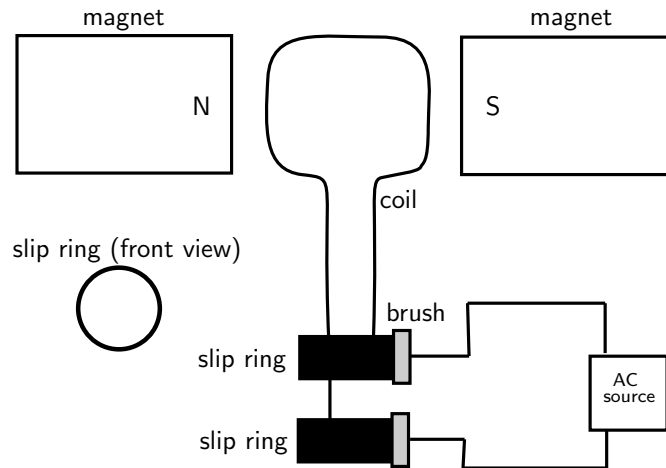


Figure 13.6: Layout of an alternating current motor.

If the current is DC, split-ring commutators are required to create a DC motor. This is shown in Figure 13.7.

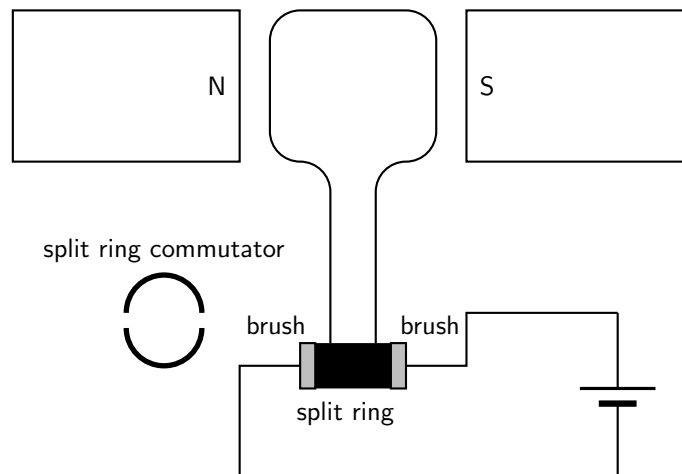


Figure 13.7: Layout of a direct current motor.

13.2.3 Real-life applications

Cars

A car contains an alternator that charges its battery and powers the car's electric system when its engine is running. Alternators have the great advantage over direct-current generators of not using a commutator, which makes them simpler, lighter, less costly, and more rugged than a DC generator.

Activity :: Research Topic : Alternators

Try to find out the different ampere values produced by alternators for different types of machines. Compare these to understand what numbers make sense in the real world. You will find different numbers for cars, trucks, buses, boats etc. Try to find out what other machines might have alternators.

A car also contains a DC electric motor, the starter motor, to turn over the engine to start it. A starter motor consists of the very powerful DC electric motor and starter solenoid that is attached to the motor. A starter motor requires a very high current to crank the engine, that's why it's connected to the battery with large cables.

Electricity Generation

AC generators are mainly used in the real-world to generate electricity.

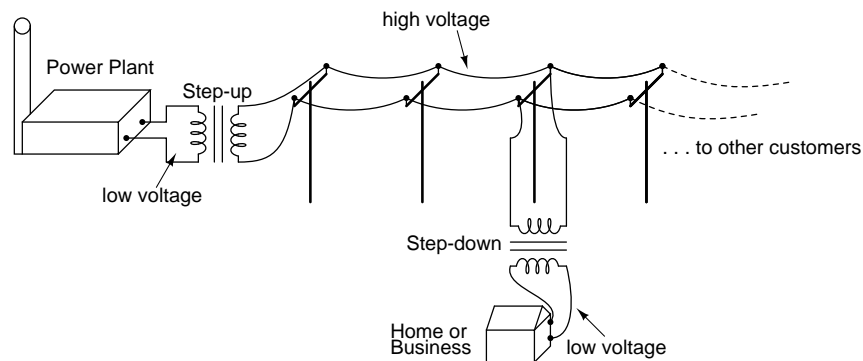


Figure 13.8: AC generators are used at the power plant to generate electricity.

13.2.4 Exercise - generators and motors

1. State the difference between a generator and a motor.
2. Use Faraday's Law to explain why a current is induced in a coil that is rotated in a magnetic field.
3. Explain the basic principle of an AC generator in which a coil is mechanically rotated in a magnetic field. Draw a diagram to support your answer.
4. Explain how a DC generator works. Draw a diagram to support your answer. Also, describe how a DC generator differs from an AC generator.
5. Explain why a current-carrying coil placed in a magnetic field (but not parallel to the field) will turn. Refer to the force exerted on moving charges by a magnetic field and the torque on the coil.
6. Explain the basic principle of an electric motor. Draw a diagram to support your answer.
7. Give examples of the use of AC and DC generators.
8. Give examples of the uses of motors.

13.3 Alternating Current

Most students learning about electricity begin with what is known as *direct current* (DC), which is electricity flowing in a constant direction. DC is the kind of electricity made by a battery, with definite positive and negative terminals.

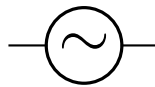
However, we have seen that the electricity produced by some generators alternates and is therefore known as *alternating current* (AC). The main advantage to AC is that the voltage can be changed

using transformers. That means that the voltage can be stepped up at power stations to a very high voltage so that electrical energy can be transmitted along power lines at low current and therefore experience low energy loss due to heating. The voltage can then be stepped down for use in buildings and street lights.



In South Africa alternating current is generated at a frequency of 50 Hz.

The circuit symbol for alternating current is:



Graphs of voltage against time and current against time for an AC circuit are shown in Figure 13.9

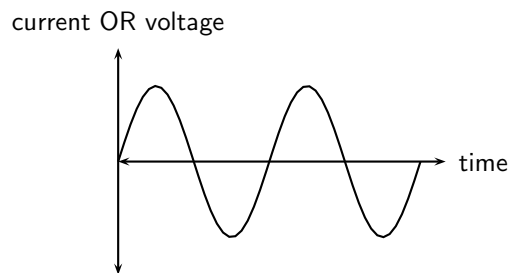


Figure 13.9: Graph of current or voltage in an AC circuit.

In an ideal DC circuit the current and voltage are constant. In an AC circuit the current and voltage vary with time. The value of the current or voltage at any specific time is called the *instantaneous current or voltage* and is calculated as follows:

$$i = I_{max} \sin(2\pi ft + \phi)$$

$$v = V_{max} \sin(2\pi ft)$$

i is the instantaneous current. I_{max} is the maximum current. v is the instantaneous voltage. V_{max} is the maximum voltage. f is the frequency of the AC and t is the time at which the instantaneous current or voltage is being calculated.

The average value we use for AC is known as the root mean square (rms) average. This is defined as:

$$I_{rms} = \frac{I_{max}}{\sqrt{2}}$$

$$V_{rms} = \frac{V_{max}}{\sqrt{2}}$$

Since AC varies sinusoidally, with as much positive as negative, doing a straight average would get you zero for the average voltage. The rms value by-passes this problem.

13.3.1 Exercise - alternating current

1. Explain the advantages of alternating current.
2. Write expressions for the current and voltage in an AC circuit.
3. Define the rms (root mean square) values for current and voltage for AC.
4. What is the period of the AC generated in South Africa?
5. If V_{max} at a power station generator is 340 V AC, what is the mains supply (rms voltage) in our household?
6. Draw a graph of voltage vs time and current vs time for an AC circuit.

13.4 Capacitance and inductance

Capacitors and inductors are found in many circuits. Capacitors store an electric field, and are used as temporary power sources as well as to minimize power fluctuations in major circuits. Inductors work in conjunction with capacitors for electrical signal processing. Here we explain the physics and applications of both.

13.4.1 Capacitance

You have learnt about capacitance and capacitors in Grade 11.

In this section you will learn about capacitance in an AC circuit. A capacitor in an AC circuit has *reactance*. Reactance in an AC circuit plays a similar role to resistance in a DC circuit. The reactance of a capacitor X_C is defined as:

$$X_C = \frac{1}{2\pi fC}$$

where C is the capacitance and f is the AC frequency.

If we examine the equation for the reactance of a capacitor, we see that the frequency is in the denominator. Therefore, when the frequency is low, the capacitive reactance is very high. This is why a capacitor blocks the flow of DC and low frequency AC because its reactance increases with decreasing frequency.

When the frequency is high, the capacitive reactance is low. This is why a capacitor allows the flow of high frequency AC because its reactance decreases with increasing frequency.

13.4.2 Inductance

Inductance (measured in henries, symbol H) is a measure of the generated emf for a unit change in current. For example, an inductor with an inductance of 1 H produces an emf of 1 V when the current through the inductor changes at the rate of $1 \text{ A}\cdot\text{s}^{-1}$.

An inductor is a passive electrical device used in electrical circuits for its property of inductance. An inductor is usually made as a coil (or solenoid) of conducting material, typically copper wire, wrapped around a core either of air or of ferromagnetic material.

Electrical current through the conductor creates a magnetic flux proportional to the current. A change in this current creates a change in magnetic flux that, in turn, generates an emf that acts to oppose this change in current.

The inductance of an inductor is determined by several factors:

- the shape of the coil; a short, fat coil has a higher inductance than one that is thin and tall.
- the material that the conductor is wrapped around.
- how the conductor is wound; winding in opposite directions will cancel out the inductance effect, and you will have only a resistor.

The inductance of a solenoid is defined by:

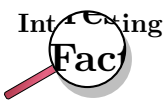
$$L = \frac{\mu_0 AN^2}{l}$$

where μ_0 is the permeability of the core material (in this case air), A is the cross-sectional area of the solenoid, N is the number of turns and l is the length of the solenoid.



Definition: Permeability

Permeability is the property of a material which describes the magnetisation developed in that material when excited by a source.



The permeability of free space is $4\pi \times 10^{-7}$ henry per metre.



Worked Example 74: Inductance I

Question: Determine the inductance of a coil with a core material of air. A cross-sectional area of $0,3\text{m}^2$, with 1000 turns and a length of $0,1$ m

Answer

Step 4 : Determine how to approach the problem

We are calculating inductance, so we use the equation:

$$L = \frac{\mu_0 AN^2}{l}$$

The permeability is that for free space: $4\pi \times 10^{-7}$ henry per metre.

Step 5 : Solve the problem

$$\begin{aligned} L &= \frac{\mu_0 AN^2}{l} \\ &= \frac{(4\pi \times 10^{-7})(0,3)(1000)}{0,1} \\ &= 3,8 \times 10^{-3} \text{H/m} \end{aligned}$$

Step 6 : Write the final answer

The inductance of the coil is $3,8 \times 10^{-3}$ H/m.



Worked Example 75: Inductance II

Question: Calculate the inductance of a 5 cm long solenoid with a diameter of 4 mm and 2000 turns.

Answer

Again this is an inductance problem, so we use the same formula as the worked example above.

$$r = \frac{4 \text{ mm}}{2} = 2 \text{ mm} = 0,002 \text{ m}$$

$$A = \pi r^2 = \pi \times 0,002^2$$

$$\begin{aligned} L &= \frac{\mu_0 AN^2}{l} \\ &= \frac{4\pi \times 10^{-7} \times 0,002^2 \times \pi \times 2000^2}{0,05} \\ &= 0,00126 \text{ H} \\ &= 1,26 \text{ mH} \end{aligned}$$

An inductor in an AC circuit also has a reactance, X_L . Reactance is the property of an inductor that opposes the flow of AC current. Reactance is defined by:

$$X_L = 2\pi fL$$

where L is the inductance and f is the frequency of the AC.

If we examine the equation for the reactance of an inductor, we see that inductive reactance increases with increasing frequency. Therefore, when the frequency is low, the inductive reactance is very low. This is why an inductor allows the flow of DC and low frequency AC because its reactance decreases with decreasing frequency.

When the frequency is high, the inductive reactance is high. This is why an inductor blocks the flow of high frequency AC because its reactance increases with increasing frequency.

13.4.3 Exercise - capacitance and inductance

1. Describe what is meant by reactance.
2. Define the reactance of a capacitor.
3. Explain how a capacitor blocks the flow of DC and low frequency AC but allows the flow of high frequency AC.
4. Describe what is an inductor.
5. Describe what is inductance.
6. What is the unit of inductance?
7. Define the reactance of an inductor.
8. Write the equation describing the inductance of a solenoid.
9. Explain how an inductor blocks high frequency AC, but allows low frequency AC and DC to pass.

13.5 Summary

1. Electrical generators convert mechanical energy into electrical energy.
2. Electric motors convert electrical energy into mechanical energy.
3. There are two types of generators - AC and DC. An AC generator is also called an alternator.
4. There are two types of motors - AC and DC.
5. Alternating current (AC) has many advantages over direct current (DC).
6. Capacitors and inductors are important components in an AC circuit.
7. The reactance of a capacitor or inductor is affected by the frequency of the AC.

13.6 End of chapter exercise

1. [SC 2003/11] Explain the difference between alternating current (AC) and direct current (DC).
2. Explain how an AC generator works. You may use sketches to support your answer.
3. What are the advantages of using an AC motor rather than a DC motor.
4. Explain how a DC motor works.
5. At what frequency is AC generated by Eskom in South Africa?

IEB 2001/11 HG1 - **Work, Energy and Power in Electric Circuits**

Mr. Smith read through the agreement with Eskom (the electricity provider). He found out that alternating current is supplied to his house at a frequency of 50 Hz. He then consulted a book on electric current, and discovered that alternating current moves to and fro in the conductor. So he refused to pay his Eskom bill on the grounds that every electron that entered his house would leave his house again, so therefore Eskom had supplied him with nothing!

Was Mr. Smith correct? Or has he misunderstood something about what he is paying for? Explain your answer briefly.

6. What do we mean by the following terms in electrodynamics?
 - (a) inductance
 - (b) reactance
 - (c) solenoid
 - (d) permeability

Chapter 14

Electronics - Grade 12

14.1 Introduction

Electronics and electrical devices surround us in daily life. From the street lights and water pumps to computers and digital phones, electronics have enabled the digital revolution to occur. All electronics are built on a backbone of simple circuits, and so an understanding of circuits is vital in understanding more complex devices.

This chapter will explain the basic physics principles of many of the components of electronic devices. We will begin with an explanation of capacitors and inductors. We see how these are used in tuning a radio. Next, we look at **active** components such as transistors and operational amplifiers. Lastly, the chapter will finish with an explanation of digital electronics, including logic gates and counting circuits.

Before studying this chapter, you will want to remind yourself of:

- The meaning of voltage (V), current (I) and resistance (R), as covered in Grade 10, and Grade 11.
- Capacitors in electric circuits, as covered in Grade 11 (see section 17.6).
- Semiconductors, as covered in Grade 11 (see chapter 20).
- The meaning of an alternating current (see section 28.3).
- Capacitance (C) and Inductance (L) (see section 28.4).

14.2 Capacitive and Inductive Circuits

Earlier in Grade 12, you were shown alternating currents (AC) and you saw that the voltage and the current varied with time. If the AC supply is connected to a resistor, then the current and voltage will be proportional to each other. This means that the current and voltage will 'peak' at the same time. We say that the current and voltage are **in phase**. This is shown in Figure 14.1.

When a capacitor is connected to an alternating voltage, the maximum voltage is proportional to the maximum current, but the maximum voltage does not occur at the same time as the maximum current. The current has its maximum (it peaks) one quarter of a cycle before the voltage peaks. Engineers say that the 'current leads the voltage by 90° '. This is shown in Figure 14.2.

For a circuit with a capacitor, the instantaneous value of $\frac{V}{I}$ is not constant. However, the value of $\frac{V_{\max}}{I_{\max}}$ is useful, and is called the **capacitive reactance** (X_C) of the component. Because it is still a voltage divided by a current (like resistance), its unit is the ohm. The value of X_C (C

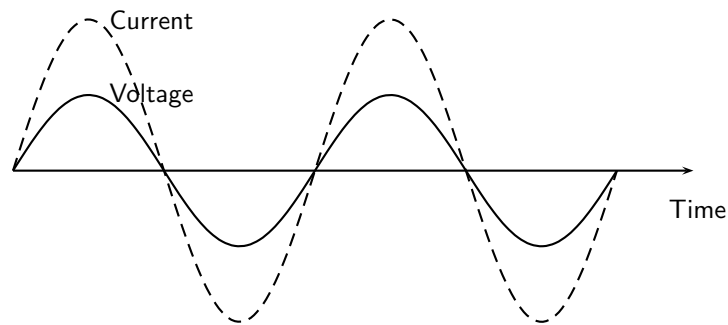


Figure 14.1: The voltage and current are in phase when a resistor is connected to an alternating voltage.

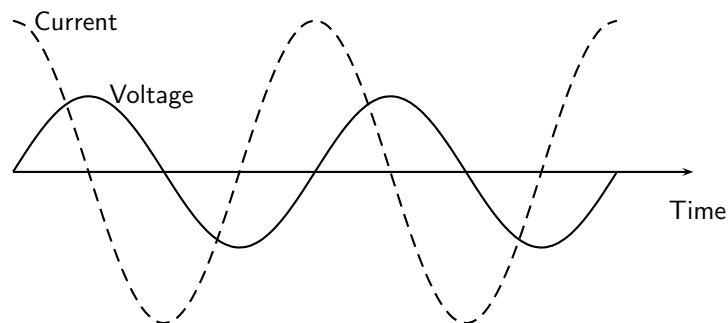


Figure 14.2: The current peaks (has its maximum) one quarter of a wave before the voltage when a capacitor is connected to an alternating voltage.

standing for capacitor) depends on its capacitance (C) and the frequency (f) of the alternating current (in South Africa 50 Hz).

$$X_C = \frac{V_{\max}}{I_{\max}} = \frac{1}{2\pi fC} \quad (14.1)$$

Inductors are very similar, but the current peaks 90° after the voltage. This is shown in Figure 14.3. Engineers say that the 'current lags the voltage'. Again, the ratio of maximum voltage to maximum current is called the reactance — this time inductive reactance (X_L). The value of the reactance depends on its inductance (L).

$$X_L = \frac{V_{\max}}{I_{\max}} = 2\pi fL \quad (14.2)$$

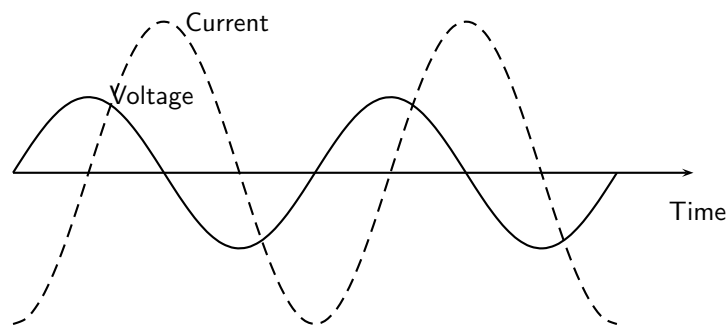


Figure 14.3: The current peaks (has its maximum) one quarter of a wave after the voltage when an inductor is connected to an alternating voltage.



Definition: Reactance

The ratio of the maximum voltage to the maximum current when a capacitor or inductor is connected to an alternating voltage. The unit of reactance is the ohm.

While inductive and capacitive reactances are similar, in one sense they are opposites. For an inductor, the current peaks 90° **after** the voltage. For a capacitor the current peaks 90° **ahead** of the voltage. When we work out the total reactance for an inductor and a capacitor in series, we use the formula

$$X_{\text{total}} = X_L - X_C \quad (14.3)$$

to take this into account. This formula can also be used when there is more than one inductor or more than one capacitor in the circuit. The total reactance is the sum of all of the inductive reactances minus the sum of all the capacitive reactances. The magnitude (number) in the final result gives the ratio of maximum voltage to maximum current in the circuit as a whole. The sign of the final result tells you its phase. If it is positive, the current peaks 90° after the voltage, if it is negative, the current peaks 90° before the voltage.

If a series circuit contains resistors as well, then the situation is more complicated. The maximum current is still proportional to the maximum voltage, but the phase difference between them won't be 90° . The ratio between the maximum voltage and maximum current is called the **impedance** (Z), and its unit is also the ohm. Impedances are calculated using this formula:

$$Z = \sqrt{X^2 + R^2} \quad (14.4)$$

where X is the total reactance of the inductors and capacitors in the circuit, and R is the total resistance of the resistors in the circuit.

It is easier to understand this formula by thinking of a right angled triangle. Resistances are drawn horizontally, reactances are drawn vertically. The hypotenuse of the triangle gives the impedance. This is shown in Figure 14.4.

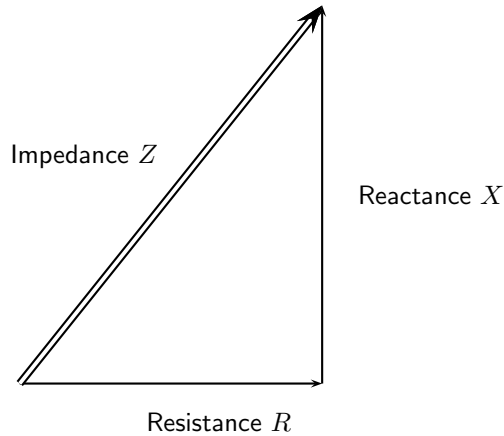


Figure 14.4: Visualizing the relationship between reactance, resistance and impedance.



Definition: Impedance

The maximum voltage divided by the maximum current for any circuit. The unit of impedance is the ohm.

It is important to remember that when resistors and inductances (or capacitors) are in a circuit, the current will not be in phase with the voltage, so the impedance is not a resistance. Similarly the current won't be exactly 90° out of phase with the voltage so the impedance isn't a reactance either.



Worked Example 76: The impedance of a coil

Question: Calculate the maximum current in a coil in a South African motor which has a resistance of 5Ω and an inductance of 3 mH . The maximum voltage across the coil is 6 V . You can assume that the resistance and inductance are in series.

Answer

1. Calculate the reactance of the coil $X_L = 2\pi fL = 2\pi \times 50 \times 0,003 = 0,942 \Omega$
2. Calculate the impedance of the coil $Z = \sqrt{X^2 + R^2} = \sqrt{0,942^2 + 5^2} = 5,09 \Omega$
3. Calculate the maximum current $I_{\max} = V_{\max}/Z = 6/5,09 = 1,18 \text{ A}$.



Worked Example 77: An RC circuit

Question: Part of a radio contains a $30\ \Omega$ resistor in series with a $3\ \mu\text{F}$ capacitor. What is its impedance at a frequency of $1\ \text{kHz}$?

Answer

1. Calculate the reactance of the capacitor

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 10^3 \times 3 \times 10^{-6}} = 53,05\ \Omega$$

2. Calculate the impedance $Z = \sqrt{X^2 + R^2} = \sqrt{53,05^2 + 30^2} = 60,9\ \Omega$

**Exercise: Capacitive and Inductive Circuits**

1. Why is the instantaneous value of $\frac{V}{I}$ of little use in an AC circuit containing an inductor or capacitor?
2. How is the reactance of an inductor different to the reactance of a capacitor?
3. Why can the ratio of the maximum voltage to the maximum current in a circuit with a resistor and an inductor not be called a reactance?
4. An engineer can describe a motor as equivalent to a $30\ \Omega$ resistor in series with a $30\ \text{mH}$ inductor. If the maximum value of the supply voltage is $350\ \text{V}$, what is the maximum current? Assume that the frequency is $50\ \text{Hz}$.
5. A timer circuit in a factory contains a $200\ \mu\text{F}$ capacitor in series with a $10\ \text{k}\Omega$ resistor. What is its impedance? Assume that the frequency is $50\ \text{Hz}$.
6. A $3\ \text{mH}$ inductor is connected in series with a $100\ \mu\text{F}$ capacitor. The reactance of the combination is zero. What is the frequency of the alternating current?



Most factories containing heavy duty electrical equipment (e.g. large motors) have to pay extra money to their electricity supply company because the inductance of the motor coils causes the current and voltage to get out of phase. As this makes the electricity distribution network less efficient, a financial penalty is incurred. The factory engineer can prevent this by connecting capacitors into the circuit to reduce the reactance to zero, as in the last question above. The current and voltage are then in phase again. We can't calculate the capacitance needed in this chapter, because the capacitors are usually connected in parallel, and we have only covered the reactances and impedances of series circuits.

14.3 Filters and Signal Tuning

14.3.1 Capacitors and Inductors as Filters

We have already seen how capacitors and inductors pass current more easily at certain frequencies than others. To recap: if we examine the equation for the reactance of a capacitor, we see that the frequency is in the denominator. Therefore, when the frequency is low, the capacitive reactance is very high. This is why a capacitor blocks the flow of DC and low frequency AC because its reactance increases with decreasing frequency.

When the frequency is high, the capacitive reactance is low. This is why a capacitor allows the flow of high frequency AC because its reactance decreases with increasing frequency.

Therefore putting a capacitor in a circuit blocks the low frequencies and allows the high frequencies to pass. This is called a high pass filter. A filter like this can be used in the 'treble' setting of a sound mixer or music player which controls the amount of high frequency signal reaching the speaker. The more high frequency signal there is, the 'tinnier' the sound. A simple high pass filter circuit is shown in Figure 14.5.

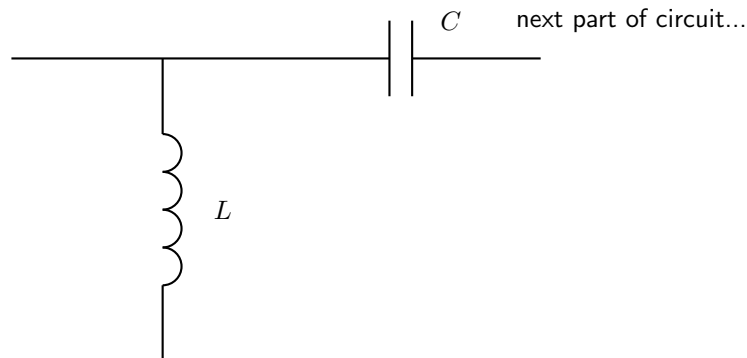


Figure 14.5: A high pass filter. High frequencies easily pass through the capacitor and into the next part of the circuit, while low frequencies pass through the inductor straight to ground.

Similarly, if we examine the equation for the reactance of an inductor, we see that inductive reactance increases with increasing frequency. Therefore, when the frequency is low, the inductive reactance is very low. This is why an inductor allows the flow of DC and low frequency AC because its reactance decreases with decreasing frequency.

When the frequency is high, the inductive reactance is high. This is why an inductor blocks the flow of high frequency AC because its reactance increases with increasing frequency.

Therefore putting an inductor in a circuit blocks the high frequencies and allows the low frequencies to pass. This is called a low pass filter. A filter like this can be used in the 'bass' setting of a sound mixer or music player which controls the amount of low frequency signal reaching the speaker. The more low frequency signal there is, the more the sound 'booms'. A simple low pass filter circuit is shown in Figure 14.6.

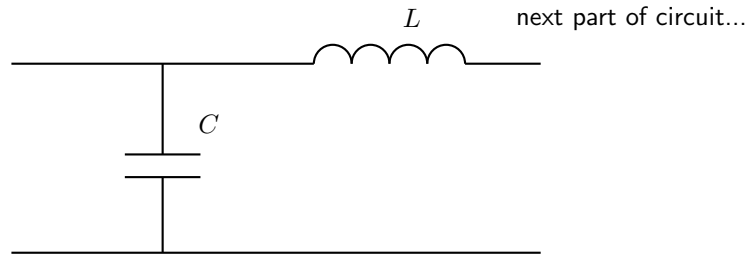


Figure 14.6: A low pass filter. Low frequencies pass through the inductor and into the next part of the circuit, while high frequencies pass through the capacitor straight to ground.

14.3.2 LRC Circuits, Resonance and Signal Tuning

A circuit containing a resistor, a capacitor and an inductor all in series is called an LRC circuit. Because the components are in series, the current through each component at a particular time will be the same as the current through the others. The voltage across the resistor will be in phase with the current. The voltage across the inductor will be 90° ahead of the current (the current always follows or lags the voltage in an inductor). The voltage across the capacitor will be 90° behind the current (the current leads the voltage for a capacitor). The phases of the voltages for the inductor and capacitor relative to the common current are shown in Figure 14.7.

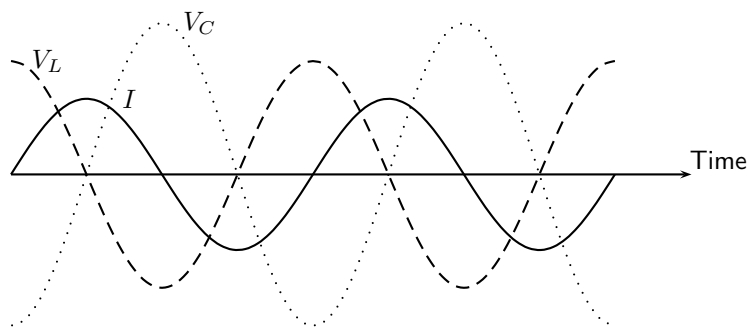


Figure 14.7: The voltages across the separate components of an LRC circuit. Looking at the peaks, you see that the voltage across the inductor V_L 'peaks' first, followed 90° later by the current I , followed 90° later by the voltage across the capacitor V_C . The voltage across the resistor is not shown — it is in phase with the current and peaks at the same time as the current.

The reactance of the inductor is $2\pi fL$, and the reactance of the capacitor is $1/2\pi fC$ but with the opposite phase. So the total reactance of the LRC circuit is

$$X = X_L - X_C = 2\pi fL - \frac{1}{2\pi fC}$$

The impedance of the circuit as a whole is given by

$$Z = \sqrt{X^2 + R^2} = \sqrt{\left(2\pi fL - \frac{1}{2\pi fC}\right)^2 + R^2}$$

At different frequencies, the impedance will take different values. The impedance will have its smallest value when the positive inductive reactance cancels out the negative capacitive reactance. This occurs when

$$2\pi fL = \frac{1}{2\pi fC}$$

so the frequency of minimum impedance must be

$$f = \frac{1}{2\pi\sqrt{LC}}$$

This is called the **resonant frequency** of the circuit. This is the frequency at which you can get the largest current for a particular supply voltage. It is also called the **natural frequency** of the circuit. This means the frequency at which the circuit would oscillate by itself.



Definition: Resonance

Resonance occurs when a circuit is connected to an alternating voltage at its natural frequency. A very large current can build up in the circuit, even with minimal power input.

An LRC circuit is very useful when we have a signal containing many different frequencies, and we only want one of them. If a mixed signal like this is connected to an LRC circuit, then only the resonant frequency (and other frequencies close to it) will drive measurable currents. This means that an LRC circuit can select one frequency from a range. This process is called **signal tuning**.



When you set up a radio antenna, and amplify the radio signal it receives, you find many different bands of frequencies — one from each radio station. When you listen to the radio, you only want to listen to one station. An LRC circuit in the radio (the tuning circuit) is set so that its resonant frequency is the frequency of the station you want to listen to. This means that of the many radio stations broadcasting, you only hear one. When you adjust the tuning dial on the radio, it changes the capacitance of the capacitor in the tuning circuit. This changes the resonant frequency, and this changes the radio station that you will hear.



Exercise: Filters and Signal Tuning

1. Which component would you use if you wanted to block low frequencies?
2. Which component would you use if you wanted to block high frequencies?
3. Calculate the impedance of a series circuit containing a $50\ \Omega$ resistor, a $30\ \mu\text{F}$ capacitor and a $3\ \text{mH}$ inductor for frequencies of (a) $50\ \text{Hz}$, (b) $500\ \text{Hz}$, and (c) $5\ 000\ \text{Hz}$.
4. Calculate the resonant frequency of the circuit in the previous question, part (c).
5. A radio station broadcasts at a frequency of $150\ \text{kHz}$. The tuning circuit in the radio contains a $0.3\ \text{mH}$ inductor. What is the capacitance of the capacitor needed in the tuning circuit if you want to listen to this radio station?
6. State the relationship between the phase of the voltages across an inductor, a resistor and a capacitor in an LRC circuit.
7. Explain what is meant by resonance.
8. Explain how LRC circuits are used for signal tuning, for example in a radio.

14.4 Active Circuit Elements

The components you have been learning about so far — resistors, capacitors and inductors — are called **passive** components. They do not change their behaviour or physics and therefore always have the same response to changes in voltage or current. **Active** components are quite different. Their response to changes in input allows them to make amplifiers, calculators and computers.

14.4.1 The Diode

A diode is an electronic device that allows current to flow in one direction only.

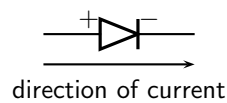


Figure 14.8: Diode circuit symbol and direction of flow of current.

A diode consists of two doped semi-conductors joined together so that the resistance is low when connected one way and very high the other way.

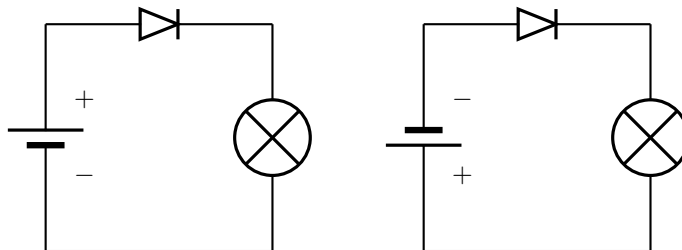


Figure 14.9: Operation of a diode. (Left) The diode is forward biased and current is permitted. The negative terminal of the battery is connected to the negative terminal of the diode. (Right) The diode is reverse biased and current flow is not allowed. The negative terminal of the battery is connected to the positive terminal of the diode.

A full explanation of diode operation is complex. Here is a simplified description. The diode consists of two semiconductor blocks attached together. Neither block is made of pure silicon — they are both **doped**. Doping was described in more detail in Section 10.3.

In short, p-type semiconductor has fewer free electrons than normal semiconductor. 'P' stands for 'positive', meaning a lack of electrons, although the material is actually neutral. The locations where electrons are missing are called **holes**. This material can conduct electricity well, because electrons can move into the holes, making a new hole somewhere else in the material. Any extra electrons introduced into this region by the circuit will fill some or all of the holes.

In n-type semiconductor, the situation is reversed. The material has more free electrons than normal semiconductor. 'N' stands for 'negative', meaning an excess of electrons, although the material is actually neutral.

When a p-type semiconductor is attached to an n-type semiconductor, some of the free electrons in the n-type move across to the p-type semiconductor. They fill the available holes near the junction. This means that the region of the n-type semiconductor nearest the junction has no free electrons (they've moved across to fill the holes). This makes this n-type semiconductor positively charged. It used to be electrically neutral, but has since lost electrons.

The region of p-type semiconductor nearest the junction now has no holes (they've been filled in by the migrating electrons). This makes the p-type semiconductor negatively charged. It used to be electrically neutral, but has since gained electrons.

Without free electrons or holes, the central region can not conduct electricity well. It has a high resistance, and is called the **depletion band**. This is shown in Figure 14.10.

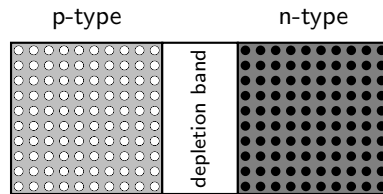


Figure 14.10: A diode consists of two doped semi-conductors joined together so that the resistance is low when connected one way and very high the other way.

You can explain the high resistance in a different way. A free electron in the n-type semiconductor will be repelled from the p-type semiconductor because of its negative charge. The electron will not go into the depletion band, and certainly won't cross the band to the p-type semiconductor. You may ask, "But won't a free electron in the p-type semiconductor be attracted across the band, carrying a current?" But there are no free electrons in p-type semiconductor, so no current of this kind can flow.

If the diode is reverse-biased, the + terminal of the battery is connected to the n-type semiconductor. This makes it even more negatively charged. It also removes even more of the free electrons near the depletion band. At the same time, the - terminal of the battery is connected to the p-type silicon. This will supply free electrons and fill in more of the holes next to the depletion band. Both processes cause the depletion band to get wider. The resistance of the diode (which was already high) increases. This is why a reverse-biased diode does not conduct.

Another explanation for the increased resistance is that the battery has made the p-type semiconductor *more negative* than it used to be, making it repel any electrons from the n-type semiconductor which attempt to cross the depletion band.

On the other hand, if the diode is forward biased, the depletion band is made narrower. The negative charge on the p-type silicon is cancelled out by the battery. The greater the voltage used, the narrower the depletion band becomes. Eventually, when the voltage is about 0,6 V (for silicon) the depletion band disappears. Once this has occurred, the diode conducts very well.



Exercise: The Diode

1. What is a diode?
2. What is a diode made of?
3. What is the term which means that a diode is connected the 'wrong way' and little current is flowing?
4. Why is a diode able to conduct electricity in one direction much more easily than the other?

14.4.2 The Light-Emitting Diode (LED)

A light-emitting diode (LED) is a diode device that emits light when charge flows in the correct direction through it. If you apply a voltage to force current to flow in the direction the LED allows, it will light up.

Light-emitting diode (LED)

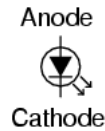


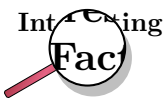
Figure 14.11: Symbol for a light-emitting diode with anode and cathode labeled.



Extension: Circuit Symbols

This notation of having two small arrows pointing away from the device is common to the schematic symbols of all light-emitting semiconductor devices. Conversely, if a device is light-activated (meaning that incoming light stimulates it), then the symbol will have two small arrows pointing toward it. It is interesting to note, though, that LEDs are capable of acting as light-sensing devices: they will generate a small voltage when exposed to light, much like a solar cell on a small scale. This property can be gainfully applied in a variety of light-sensing circuits.

The colour depends on the semiconducting material used to construct the LED, and can be in the near-ultraviolet, visible or infrared part of the electromagnetic spectrum.



Nick Holonyak Jr. (1928) of the University of Illinois at Urbana-Champaign developed the first practical visible-spectrum LED in 1962.

Light emission

The wavelength of the light emitted, and therefore its colour, depends on the materials forming the p-n junction. A normal diode, typically made of silicon or germanium, emits invisible far-infrared light (so it can't be seen), but the materials used for an LED can emit light corresponding to near-infrared, visible or near-ultraviolet frequencies.

LED applications

LEDs have many uses. Some of these are given here.

- thin, lightweight message displays, e.g. in public information signs (at airports and railway stations, among other places)
- status indicators, e.g. on/off lights on professional instruments and consumers audio/video equipment

- infrared LEDs in remote controls (for TVs, VCRs, etc.)
- clusters of LEDs are used in traffic signals, replacing ordinary bulbs behind coloured glass
- car indicator lights and bicycle lighting
- calculator and measurement instrument displays (seven segment displays), although now mostly replaced by LCDs
- red or yellow LEDs are used in indicator and [alpha]numeric displays in environments where night vision must be retained: aircraft cockpits, submarine and ship bridges, astronomy observatories, and in the field, e.g. night time animal watching and military field use
- red or yellow LEDs are also used in photographic darkrooms, for providing lighting which does not lead to unwanted exposure of the film
- illumination, e.g. flashlights (a.k.a. torches, UK), and backlighting for LCD screens
- signaling/emergency beacons and strobes
- movement sensors, e.g. in mechanical and optical computer mice and trackballs
- in LED printers, e.g. high-end colour printers

LEDs offer benefits in terms of maintenance and safety.

- The typical working lifetime of a device, including the bulb, is ten years, which is much longer than the lifetimes of most other light sources.
- LEDs fail by dimming over time, rather than the abrupt burn-out of incandescent bulbs.
- LEDs give off less heat than incandescent light bulbs and are less fragile than fluorescent lamps.
- Since an individual device is smaller than a centimetre in length, LED-based light sources used for illumination and outdoor signals are built using clusters of tens of devices.

Because they are monochromatic, LED lights have great power advantages over white lights where a specific colour is required. Unlike the white lights, the LED does not need a filter that absorbs most of the emitted white light. Coloured fluorescent lights are made, but they are not widely available. LED lights are inherently coloured, and are available in a wide range of colours. One of the most recently introduced colours is the emerald green (bluish green, wavelength of about 500 nm) that meets the legal requirements for traffic signals and navigation lights.



The largest LED display in the world is 36 m high, at Times Square, New York, U.S.A.

There are applications that specifically require light that does not contain any blue component. Examples are photographic darkroom safe lights, illumination in laboratories where certain photo-sensitive chemicals are used, and situations where dark adaptation (night vision) must be preserved, such as cockpit and bridge illumination, observatories, etc. Yellow LED lights are a good choice to meet these special requirements because the human eye is more sensitive to yellow light.



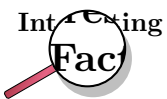
Exercise: The Light Emitting Diode

1. What is an LED?
 2. List 5 applications of LEDs.
-

14.4.3 Transistor

The diode is the simplest semiconductor device, made up of a p-type semiconductor and an n-type semiconductor in contact. It can conduct in only one direction, but it cannot control the size of an electric current. Transistors are more complicated electronic components which can control the size of the electric current flowing through them.

This enables them to be used in amplifiers. A small signal from a microphone or a radio antenna can be used to control the transistor. In response, the transistor will then increase and decrease a much larger current which flows through the speakers.



One of the earliest popular uses of transistors was in cheap and portable radios. Before that, radios were much more expensive and contained glass valves which were fragile and needed replacing. In some parts of the world you can still hear people talking about their 'transistor' — they mean their portable radio.

You can also use a small current to turn the transistor on and off. The transistor then controls a more complicated or powerful current through other components. When a transistor is used in this way it is said to be in **switching** mode as it is acting as a remotely controlled switch. As we shall see in the final sections of this chapter, switching circuits can be used in a computer to process and store digital information. A computer would not work without the millions (or billions) of transistors in it.

There are two main types of transistor - bipolar transistors (NPN or PNP), and field effect transistors (FETs). Both use doped semiconductors, but in different ways. You are mainly required to know about field effect transistors (FETs), however we have to give a brief description of bipolar transistors so that you see the difference.

Bipolar Transistors

Bipolar transistors are made of a doped semiconductor 'sandwich'. In an NPN transistor, a very thin layer of p-type semiconductor is in between two thicker layers of n-type semiconductor. This is shown in Figure 14.12. Similarly an PNP transistor consists of a very thin n-type layer in between two thicker layers of p-type semiconductor.

In an NPN transistor a small current of electrons flows from the emitter (E) to the base (B). Simultaneously, a much larger current of electrons flows from the emitter (E) to the collector (C). If you lower the number of electrons able to leave the transistor at the base (B), the transistor automatically reduces the number of electrons flowing from emitter (E) to collector (C). Similarly, if you increase the current of electrons flowing out of the base (B), the transistor automatically also increases the current of electrons flowing from emitter (E) to collector (C). The transistor is designed so that the current of electrons from emitter to collector (I_{EC}) is proportional to

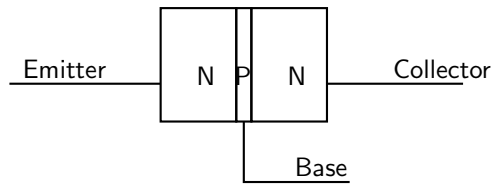


Figure 14.12: An NPN transistor. This is a type of bipolar transistor.

the current of electrons from emitter to base (I_{EB}). The constant of proportionality is known as the **current gain** β . So $I_{EC} = \beta I_{EB}$.

How does it do it? The answer comes from our work with diodes. Electrons arriving at the emitter (n-type semiconductor) will naturally flow through into the central p-type since the base-emitter junction is forward biased. However if none of these electrons are removed from the base, the electrons flowing into the base from the emitter will fill all of the available 'holes'. Accordingly, a large depletion band will be set up. This will act as an insulator preventing current flow into the collector as well. On the other hand, if the base is connected to a positive voltage, a small number of electrons will be removed by the base connection. This will prevent the 'holes' in the base becoming filled up, and no depletion band will form. While some electrons from the emitter leave via the base connection, the bulk of them flow straight on to the collector. You may wonder how the electrons get from the base into the collector (it seems to be reverse biased). The answer is complicated, but the important fact is that the p-type layer is extremely thin. As long as there is no depletion layer, the bulk of the electrons will have no difficulty passing straight from the n-type emitter into the n-type collector. A more satisfactory answer can be given to a university student once band theory has been explained.

Summing up, in an NPN transistor, a small flow of electrons from emitter (E) to base (B) allows a much larger flow of electrons from emitter (E) to collector (C). Given that conventional current (flowing from + to -) is in the opposite direction to electron flow, we say that a small conventional current from base to emitter allows a large current to flow from collector to emitter.

A PNP transistor works the other way. A small conventional current from emitter to base allows a much larger conventional current to flow from emitter to collector. The operation is more complicated to explain since the principal charge carrier in a PNP transistor is not the electron but the 'hole'.

The operation of NPN and PNP transistors (in terms of conventional currents) is summarized in Figure 14.13.



The transistor is considered by many to be one of the greatest discoveries or inventions in modern history, ranking with banking and the printing press. Key to the importance of the transistor in modern society is its ability to be produced in huge numbers using simple techniques, resulting in vanishingly small prices. Computer "chips" consist of millions of transistors and sell for Rands, with per-transistor costs in the thousandths-of-cents. The low cost has meant that the transistor has become an almost universal tool for non-mechanical tasks. Whereas a common device, say a refrigerator, would have used a mechanical device for control, today it is often less expensive to simply use a few million transistors and the appropriate computer program to carry out the same task through "brute force". Today transistors have replaced almost all electromechanical devices, most simple feedback systems, and appear in huge numbers in everything from computers to cars.

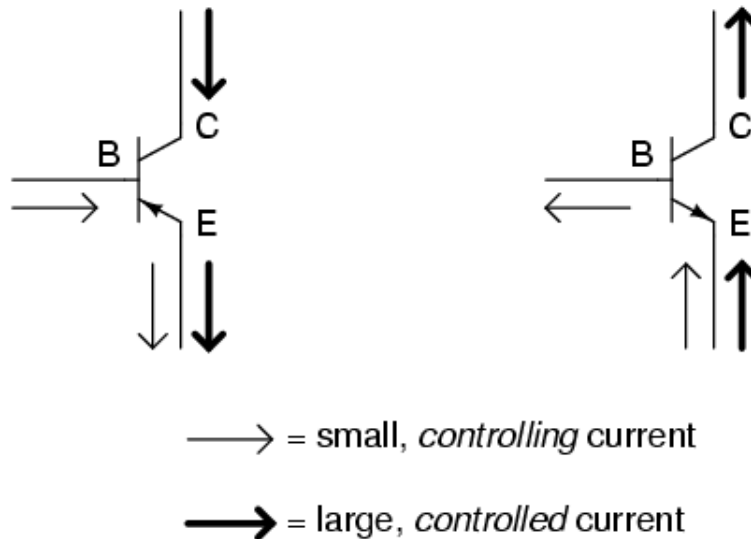


Figure 14.13: An overview of bipolar transistors as current amplifiers. (Left) An NPN transistor. (Right) A PNP transistor.



The transistor was invented at Bell Laboratories in December 1947 (first demonstrated on December 23) by John Bardeen, Walter Houser Brattain, and William Bradford Shockley, who were awarded the Nobel Prize in physics in 1956.

The Field Effect Transistor (FET)

To control a bipolar transistor, you control the **current** flowing into or out of its base. The other type of transistor is the field effect transistor (FET). FETs work using control **voltages** instead. Accordingly they can be controlled with much smaller currents and are much more economic to use.



No-one would build a computer with billions of bipolar transistors — the current in each transistor's base might be small, but when you add up all of the base currents in the millions of transistors, the computer as a whole would be consuming a great deal of electricity and making a great deal of heat. Not only is this wasteful, it would prevent manufacturers making a computer of convenient size. If the transistors were too close together, they would overheat.

The three terminals of the FET are called the *source* (S), *drain* (D) and *gate* (G), as shown in Figure 14.14. When the gate is not connected, a current of electrons can flow from source (S) to drain (D) easily along the channel. The source is, accordingly, the negative terminal of the transistor. The drain, where the electrons come out, is the positive terminal of the transistor. A few electrons will flow from the n-type channel into the p-type semiconductor of the gate

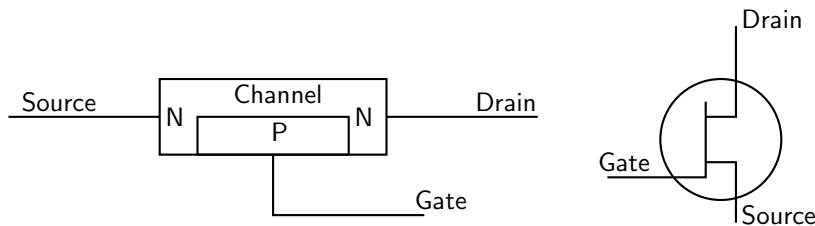


Figure 14.14: A field effect transistor (FET). The diagram on the left shows the semiconductor structure. The diagram on the right shows its circuit symbol.

when the device is manufactured. However, as these electrons are not removed (the gate is not connected), a depletion band is set up which prevents further flow into the gate.

In operation, the gate is connected to negative voltages relative to the source. This makes the p-n junction between gate and channel reverse-biased. Accordingly no current flows from the source into the gate. When the voltage of the gate is lowered (made more negative), the depletion band becomes wider. This enlarged depletion band takes up some of the space of the channel. So the lower the voltage of the gate (the more negative it is relative to the source), the larger the depletion band. The larger the depletion band, the narrower the channel. The narrower the channel, the harder it is for electrons to flow from source to drain.

The voltage of the gate is not the only factor affecting the current of electrons between the source and the drain. If the external circuit has a low resistance, electrons are able to leave the drain easily. If the external circuit has a high resistance, electrons leave the drain slowly. This creates a kind of 'traffic jam' which slows the passage of further electrons. In this way, the voltage of the drain regulates itself, and is more or less independent of the current demanded from the drain.

Once these two factors have been taken into account, it is fair to say that the positive output voltage (the voltage of the drain relative to the source) is proportional to the negative input voltage (the voltage of the gate relative to the source).

For this reason, the field effect transistor is known as a voltage amplifier. This contrasts with the bipolar transistor which is a current amplifier.



Exercise: Field Effect Transistors

1. What are the two types of bipolar transistor? How does their construction differ?
2. What are the three connections to a bipolar transistor called?
3. Why are very few electrons able to flow from emitter to collector in an NPN transistor if the base is not connected?
4. Why do you think a bipolar transistor would not work if the base layer were too thick?
5. "The bipolar transistor is a current amplifier." What does this statement mean?
6. Describe the structure of a FET.
7. Define what is meant by the source, drain and gate. During normal operation, what will the voltages of drain and gate be with respect to the source?
8. Describe how a depletion layer forms when the gate voltage is made more negative. What controls the width of the depletion layer?
9. "The field effect transistor is a voltage amplifier." What does this statement mean?

10. The amplifier in a cheap radio will probably contain bipolar transistors. A computer contains many field effect transistors. Bipolar transistors are more rugged and less sensitive to interference than field effect transistors, which makes them more suitable for a simple radio. Why are FETs preferred for the computer?

14.4.4 The Operational Amplifier

The operational amplifier is a special kind of voltage amplifier which is made from a handful of bipolar or field effect transistors. Operational amplifiers are usually called **op-amps** for short. They are used extensively in all kinds of audio equipment (amplifiers, mixers and so on) and in instrumentation. They also have many other uses in other circuits - for example comparing voltages from sensors.

Operational amplifiers are supplied on Integrated Circuits (I.C.s). The most famous operational amplifier I.C. is numbered 741 and contains a single operational amplifier on an integrated circuit ('chip') with eight terminals. Other varieties can be bought, and you can get a single integrated circuit with two or four '741'-type operational amplifiers on it.

The symbol for an op-amp is shown in Figure 14.15. The operational amplifier has two input terminals and one output terminal. The voltage of the output terminal is proportional to the difference in voltage between the two input terminals. The output terminal is on the right (at the sharp point of the triangle). The two input terminals are drawn on the left. One input terminal (labelled with a + on diagrams) is called the **non-inverting input**. The other input terminal (labelled -) is called the **inverting input**. The labels + and - have nothing to do with the way in which the operational amplifier is connected to the power supply. Operational amplifiers must be connected to the power supply, but this is taken for granted when circuit diagrams are drawn, and these connections are not shown on circuit diagrams. Usually, when drawing electronic circuits, '0V' is taken to mean the negative terminal of the power supply. This is not the case with op-amps. For an op-amp, '0V' refers to the voltage midway between the + and - of the supply.

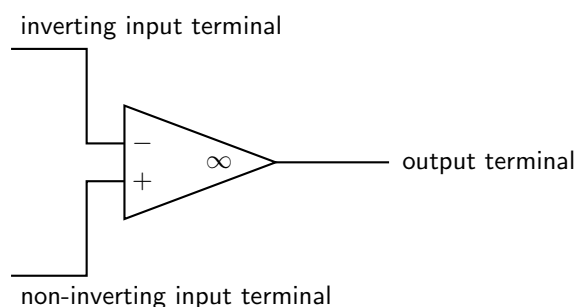


Figure 14.15: Circuit symbol for an operational amplifier. The amplifier must also be connected to the + and - terminals of the power supply. These connections are taken for granted and not shown.

The output voltage of the amplifier V_{out} is given by the formula

$$V_{out} = A(V_+ - V_-) \quad (14.5)$$

here A is a constant called the **open loop gain**, and V_+ and V_- are the voltages of the two input terminals. That said, the output voltage can not be less than the voltage of the negative terminal of the battery supplying it or higher than the positive terminal of the battery supplying

it. You will notice that V_{out} is positive if $V_+ > V_-$ and negative if $V_+ < V_-$. This is why the $-$ input is called the inverting input: raising its voltage causes the output voltage to *drop*.

The input resistance of an operational amplifier is very high. This means that very little current flows into the input terminals during operation.

If all of the transistors in the operational amplifier were identical then the output voltage would be zero if the two inputs were at equal voltages. In practice this is not quite the case, and for sensitive work a **trimming potentiometer** is connected. This is adjusted until the op-amp is zeroed correctly.

Simple operational amplifiers require the trimming potentiometer to be built into the circuit containing them, and an example is shown in Figure 14.16. Other operational amplifier designs incorporate separate terminals for the trimming potentiometer. These special terminals are labelled **offset** on the manufacturer's diagram. The exact method of connecting the potentiometer to the offset terminals can depend on the design of the operational amplifier, and you need to refer to the manufacturer's data sheet for details of which potentiometer to use and how to connect it.

For most commercially produced operational amplifiers (known as op-amps for short), the open loop gain A is very large and does not stay constant. Values of 100 000 are typical. Usually a designer would want an amplifier with a stable gain of smaller value, and builds the operational amplifier into a circuit like the one in Figure 14.16.

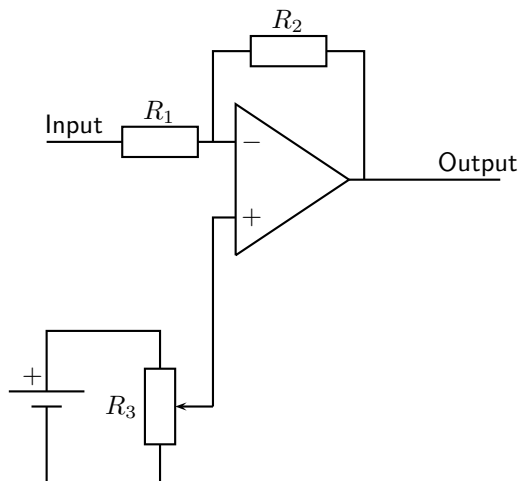


Figure 14.16: An inverting amplifier built using an operational amplifier. The connections from battery to operational amplifier are not shown. The output voltage $V_{out} = -R_2V_{in}/R_1$, as explained in the text. The potentiometer R_3 is a trimming potentiometer. To set it, the input is connected to zero volts. The trimming potentiometer is then adjusted until $V_{out} = 0$. In all operational amplifier circuits, zero volts is midway between the $+$ and $-$ of the supply.



Extension: Calculating the gain of the amplifier in Figure 14.16.

1. The input resistance of the operational amplifier is very high. This means that very little current flows into the inverting input of the op-amp. Accordingly, the current through resistor R_1 must be almost the same as the current through resistor R_2 . This means that the ratio of the voltage across R_1 to the voltage across R_2 is the same as the ratio of the two resistances.
2. The open loop gain A of the op-amp is very high. Assuming that the output voltage is less than a few volts, this means that the two input terminals must be at very similar voltages. We shall assume that they are at the same voltage.
3. We want the output voltage to be zero if the input voltage is zero. Assuming that the transistors within the op-amp are very similar, the output voltage will

only be zero for zero input voltage if V_+ is very close to zero. We shall assume that $V_+ = 0$ when the trimming potentiometer is correctly adjusted.

4. It follows from the last two statements that $V_- \approx 0$, and we shall assume that it is zero.
5. With these assumptions, the voltage across R_2 is the same as V_{out} , and the voltage across R_1 is the same as V_{in} . Since both resistors carry the same current (as noted in point 1), we may say that the magnitude of $V_{out}/V_{in} = R_2/R_1$. However, if V_{in} is negative, then V_{out} will be positive. Therefore it is customary to write the gain of this circuit as $V_{out}/V_{in} = -R_2/R_1$.



Exercise: Operational Amplifiers

1. What are operational amplifiers used for?
 2. Draw a simple diagram of an operational amplifier and label its terminals.
 3. Why is a trimming potentiometer needed when using an op-amp?
-

14.5 The Principles of Digital Electronics

The circuits and components we have discussed are very useful. You can build a radio or television with them. You can make a telephone. Even if that was all there was to electronics, it would still be very useful. However, the great breakthrough in the last fifty years or so has been in **digital** electronics. This is the subject which gave us the computer. The computer has revolutionized the way business, engineering and science are done. Small computers programmed to do a specific job (called microprocessors) are now used in almost every electronic machine from cars to washing machines. Computers have also changed the way we communicate. We used to have telegraph or telephone wires passing up and down a country — each one carrying one telephone call or signal. We now have optic fibres each capable of carrying tens of thousands of telephone calls using **digital** signals.

So, what is a digital signal? Look at Figure 14.17. A normal signal, called an **analogue** signal, carries a smooth wave. At any time, the voltage of the signal could take any value. It could be 2,00 V or 3,53 V or anything else. A digital signal can only take certain voltages. The simplest case is shown in the figure — the voltage takes one of two values. It is either **high**, or it is **low**. It never has any other value.

These two special voltages are given symbols. The low voltage level is written 0, while the high voltage level is written as 1. When you send a digital signal, you set the voltage you want (0 or 1), then keep this fixed for a fixed amount of time (for example 0.01 μ s), then you send the next 0 or 1. The digital signal in Figure 14.17 could be written 01100101.

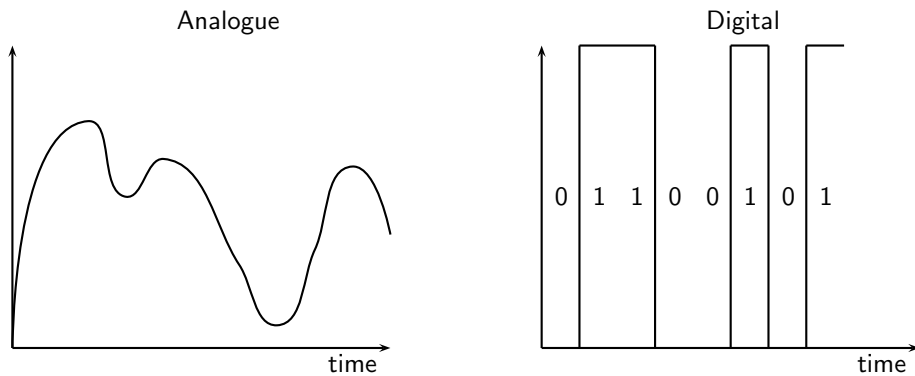


Figure 14.17: The difference between normal (analogue) signals and digital signals.

Why are digital signals so good?

1. Using a computer, any information can be turned into a pattern of 0s and 1s. Pictures, recorded music, text and motion pictures can all be turned into a string of 0s and 1s and transmitted or stored in the same way. The computer receiving the signal at the other end converts it back again. A Compact Disc (CD) for example, can store music or text or pictures, and all can be read using a computer.
2. The 0 and the 1 look very different. You can immediately tell if a 0 or a 1 is being sent. Even if there is interference, you can still tell whether the sender sent a 0 or a 1. This means that fewer mistakes are made when reading a digital signal. This is why the best music recording technologies, and the most modern cameras, for example, all use digital technology.
3. Using the 0s and 1s you can count, and do all kinds of mathematics. This will be explained in more detail in the next section.

The simplest digital circuits are called **logic gates**. Each logic gate makes a decision based on the information it receives. Different logic gates are set up to make the decisions in different ways. Each logic gate will be made of many microscopic transistors connected together within a thin wafer of silicon. This tiny circuit is called an Integrated Circuit or I.C. - all the parts are in one place (integrated) on the silicon wafer.

14.5.1 Logic Gates

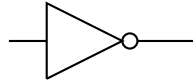
There are five main types of logic gate: NOT, AND, OR, NAND and NOR. Each one makes its decision in a different way.

The NOT Gate

Problem: You want an automatic circuit in your office to turn on the heating in the winter. You already have a digital electronic temperature sensor. When the temperature is high, it sends out a 1. When the office is cold, it sends out a 0. If this signal were sent straight to the heater, the heater would turn on (1) when it was already hot, and would stay off when it was cold. This is wrong! To make the heater work, we need a circuit which will change a 0 (from the sensor) into a 1 (to send to the heater). This will make the heater come on when it is cold. You also want it to change a 1 (from the sensor) into a 0 (to send to the heater). This will turn the heater off

when the room is hot. This circuit is called an **inverter** or **NOT gate**. It changes 0 into 1 (1 is NOT 0). It changes 1 into 0 (0 is NOT 1). It changes a signal into what it is NOT.

The symbol for the NOT gate is:



The action of the NOT gate can be written in a table called a **truth table**. The left column shows the possible inputs on different rows. The right column shows what the output (decision) of the circuit will be for that input. The truth table for the NOT gate is shown below.

Input	Output
0	1
1	0

When you read the truth table, the top row says, "If the input is 0, the output will be 1." For our heater, this means, "If the room is cold, the heater will turn on." The bottom row says, "If the input is 1, the output will be 0." For our heater, this means, "If the room is hot, the heater will switch off."

The AND Gate

Problem: An airliner has two toilets. Passengers get annoyed if they get up from their seat only to find that both toilets are being used and they have to go back to their seat and wait. You want to fit an automatic circuit to light up a display if both toilets are in use. Then passengers know that if the light is off, there will be a free toilet for them to use. There is a sensor in each toilet. It gives out a 0 if the toilet is free, and a 1 if it is in use. You want to send a 1 to the display unit if **both** sensors are sending 1s. To do this, you use an AND gate.

The symbol for the AND gate is:

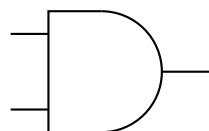


Figure 14.18: Symbol for the AND logic gate.

The truth table for the AND gate is shown below. An AND gate has two inputs (the NOT gate only had one). This means we need four rows in the truth table, one for each possible set of inputs. The first row, for example, tells us what the AND gate will do if both inputs are 0. In our airliner, this means that both toilets are free. The right column has a 0 showing that the output will be 0, so the display will not light up. The second row has inputs of 0 and 1 (the first toilet is free, the other is in use). Again the output is 0. The third row tells us what will happen if the inputs are 1 and 0 (the first toilet is in use, and the second is free). Finally, the last line tells us what will happen if both inputs are 1 (the first toilet is in use and the second toilet is in use). It is only in this case that the output is 1 and the display lights up.

Inputs		Output
A	B	
0	0	0
0	1	0
1	0	0
1	1	1

This device is called an AND gate, because the output is only 1 if one input AND the other input are both 1.



Extension: Using 0 and 1 to mean True and False

When we use logic gates we use the low voltage state 0 to represent 'false'. The high voltage state 1 represents 'true'. This is why the word AND is so appropriate. A AND B is true (1) if, and only if, A is true (1) AND B is true (1).



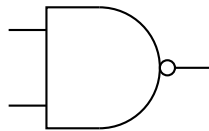
Extension: AND and multiplication

Sometimes, the AND operation is written as multiplication. A AND B is written AB. If either A or B are 0, then AB will also be 0. For AB to be 1, we need A and B to both be 1. Multiplication of the numbers 0 and 1 does exactly the same job as an AND gate.

The NAND Gate

Problem: You build the circuit for the airliner toilets using an AND gate. Your customer is pleased, but she says that it would be better if the display lit up when there **was** a free toilet. In other words, the display should light up unless both toilets are in use. To do this we want a circuit which does the opposite of an AND gate. We want a circuit which would give the output 0 where an AND gate would give 1. We want a circuit which would give the output 1 where an AND gate would give 0. This circuit is called a NAND gate.

The symbol for the NAND gate is:



The truth table for the NAND gate is shown below.

Inputs		Output
A	B	
0	0	1
0	1	1
1	0	1
1	1	0

You may have noticed that we could have done this job on the airliner by using our earlier circuit, with a NOT gate added between the original AND gate and the display. This is where the word NAND comes from — it is short for NotAND.

The OR Gate

Problem: A long, dark corridor has two light switches — one at each end of the corridor. The switches each send an output of 0 to the control unit if no-one has pressed the switch. If someone presses the switch, its output is 1. The lights in the corridor should come on if either switch is pressed. To do this job, the control unit needs an OR gate. The symbol for the OR gate is:

H

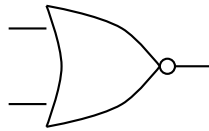


Figure 14.20: Symbol for the NOR logic gate.

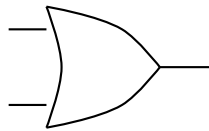


Figure 14.19: Symbol for the OR logic gate.

The truth table for the OR gate is shown.

Inputs		Output
A	B	
0	0	0
0	1	1
1	0	1
1	1	1

You can see that the output is 1 (and the lights come on in the corridor) if either one switch OR the other is pressed. Pressing both switches also turns on the lights, as the last row in the table shows.



Extension: OR and addition

Sometimes you will see $A \text{ OR } B$ written mathematically as $A+B$. This makes sense, since if $A=0$ and $B=0$, then $A \text{ OR } B = A+B = 0$. Similarly, if $A=0$ and $B=1$, then $A \text{ OR } B = A+B = 1$. If $A=1$ and $B=0$, then $A \text{ OR } B = A+B = 1$ once again. The only case where the OR function differs from normal addition is when $A=1$ and $B=1$. Here $A \text{ OR } B = 1$ in logic, but $A+B=2$ in arithmetic. However, there is no such thing as '2' in logic, so we define $+$ to mean 'OR', and write $1+1=1$ with impunity!

If you wish, you can prove that the normal rules of algebra still work using this notation: $A+(B+C) = (A+B)+C$, $A(BC) = (AB)C$, and $A(B+C) = AB + AC$. This special kind of algebra where variables can only be 0 (representing false) or 1 (representing true) is called Boolean algebra.

The NOR Gate

The last gate you need to know is the NOR gate. This is opposite to the OR gate. The output is 1 if both inputs are 0. In other words, the output switches on if neither the first NOR the second input is 1. The symbol for the NOR gate is:

The truth table for the NOR gate is shown below.

Inputs		Output
A	B	
0	0	1
0	1	0
1	0	0
1	1	0

The examples given were easy. Each job only needed one logic gate. However any 'decision making' circuit can be built with logic gates, no matter how complicated the decision. Here is an example.

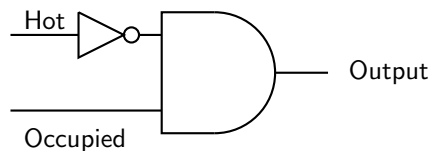


Worked Example 78: An Economic Heating Control

Question: A sensor in a building detects whether a room is being used. If it is empty, the output is 0, if it is in use, the output is 1. Another sensor measures the temperature of the room. If it is cold, the output is 0. If it is hot, the output is 1. The heating comes on if it receives a 1. Design a control circuit so that the heating only comes on if the room is in use and it is cold.

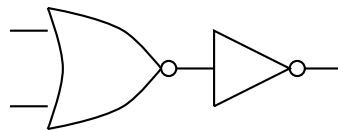
Answer

The first sensor tells us whether the room is occupied. The second sensor tells us whether the room is hot. The heating must come on if the room is occupied AND cold. This means that the heating should come on if the room is occupied AND (NOT hot). To build the circuit, we first attach a NOT gate to the output of the temperature sensor. This output of the NOT gate will be 1 only if the room is cold. We then attach this output to an AND gate, together with the output from the other sensor. The output of the AND gate will only be 1 if the room is occupied AND the output of the NOT gate is also 1. So the heating will only come on if the room is in use and is cold. The circuit is shown below.



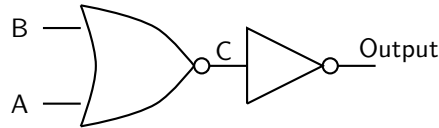
Worked Example 79: Solving a circuit with two logic gates

Question: Compile the truth table for the circuit below.



Answer

Firstly, we label the inputs A and B. We also label the point where the two gates are connected C.



Next we prepare a truth table. There is a column for each of the inputs, for the intermediate point C and also for the output. The truth table has four rows, since there are four possible inputs — 00, 01, 10 and 11.

A	B	C	Output
0	0		
0	1		
1	0		
1	1		

Next we fill in the C column given that we know what a NOR gate does.

A	B	C	Output
0	0	1	
0	1	0	
1	0	0	
1	1	0	

Next, we can fill in the output, since it will always be the opposite of C (because of the NOT gate).

A	B	C	Output
0	0	1	0
0	1	0	1
1	0	0	1
1	1	0	1

Finally we see that this combination of gates does the same job as an OR gate.

Each logic gate is manufactured from two or more transistors. Other circuits can be made using logic gates, as we shall see in the next section. We shall show you how to count and store numbers using logic gates. This means that if you have enough transistors, and you connect them correctly to make the right logic gates, you can make circuits which count and store numbers.

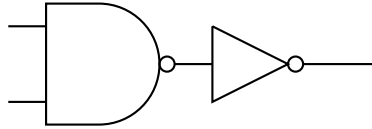
In practice, the cheapest gate to manufacture is usually the NAND gate. Additionally, Charles Peirce showed that NAND gates alone (as well as NOR gates alone) can be used to reproduce all the other logic gates.



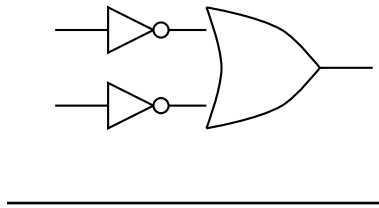
Exercise: The Principles of Digital Electronics

1. Why is digital electronics important to modern technology and information processing?
2. What two symbols are used in digital electronics, to represent a “high” and a “low”? What is this system known as?
3. What is a logic gate?

4. What are the five main types of logic gates? Draw the symbol for each logic gate.
5. Write out the truth tables for each of the five logic gates.
6. Write out the truth table for the following circuit. Which single gate is this circuit equivalent to?



7. Write out the truth table for the following circuit. Which single gate is this circuit equivalent to?



14.6 Using and Storing Binary Numbers

In the previous section, we saw how the numbers 0 and 1 could represent 'false' and 'true' and could be used in decision making. Often we want to program a computer to count with numbers. To do this we need a way of writing any number using nothing other than 0 and 1. When written in this way, numbers are called **binary numbers**.



Definition: Binary Number System

A way of writing any number using only the digits 0 and 1.

14.6.1 Binary numbers

In normal (denary) numbers, we write $9+1$ as 10. The fact that the '1' in 10 is the second digit from the right tells us that it actually means 10 and not 1. Similarly, the '3' in 365 represents 300 because it is the third digit from the right. You could write 365 as $3 \times 100 + 6 \times 10 + 5$. You will notice the pattern that the n th digit from the right represents 10^{n-1} . In binary, we use the n th digit from the right to represent 2^{n-1} . Thus 2 is written as 10 in binary. Similarly $2^2 = 4$ is written as 100 in binary, and $2^3 = 8$ is written as 1000 in binary.



Worked Example 80: Conversion of Binary Numbers to Denary Numbers

Question: Convert the binary number 10101 to its denary equivalent.

Answer

We start on the right. The '1' on the right does indeed represent one. The next '1' is in the third place from the right, and represents $2^2 = 4$. The next '1' is in the fifth place from the right and represents $2^4 = 16$. Accordingly, the binary number 10101 represents $16+4+1 = 21$ in denary notation.



Worked Example 81: Conversion of Denary Numbers to Binary Numbers

Question: Convert the decimal number 12 to its binary equivalent.

Answer

Firstly we write 12 as a sum of powers of 2, so $12 = 8 + 4$. In binary, eight is 1000, and four is 100. This means that twelve = eight + four must be $1000 + 100 = 1100$ in binary. You could also write 12 as $1 \times 8 + 1 \times 4 + 0 \times 2 + 0 \times 1 = 1100$ in binary.



How do you write numbers as a sum of powers of two? The first power of two (the largest) is the largest power of two which is not larger than the number you are working with. In our last example, where we wanted to know what twelve was in binary, the largest power of two which is not larger than 12 is 8. Thus $12 = 8 + \text{something}$. By arithmetic, the 'something' must be 4, and the largest power of two not larger than this is 4 exactly. Thus $12 = 8 + 4$, and we have finished.

A more complicated example would be to write one hundred in binary. The largest power of two not larger than 100 is 64 (1000000 in binary). Subtracting 64 from 100 leaves 36. The largest power of two not larger than 36 is 32 (100000 in binary). Removing this leaves a remainder of 4, which is a power of two itself (100 in binary). Thus one hundred is $64 + 32 + 4$, or in binary $1000000 + 100000 + 100 = 1100100$.

Once a number is written in binary, it can be represented using the low and high voltage levels of digital electronics. We demonstrate how this is done by showing you how an electronic counter works.

14.6.2 Counting circuits

To make a counter you need several 'T flip flops', sometimes called 'divide by two' circuits. A T flip flop is a digital circuit which swaps its output (from 0 to 1 or from 1 to 0) whenever the input changes from 1 to 0. When the input changes from 0 to 1 it doesn't change its output. It is called a **flip flop** because it changes (flips or flops) each time it receives a pulse.

If you put a series of pulses 10101010 into a T flip flop, the result is 01100110. Figure 14.21 makes this clearer.

As you can see from Figure 14.21, there are half as many pulses in the output. This is why it is called a 'divide by two' circuit.

If we connect T flip flops in a chain, then we make a counter which can count pulses. As an example, we connect three T flip flops in a chain. This is shown in Figure 14.22.

When this circuit is fed with a stream of pulses, the outputs of the different stages change. The table below shows how this happens. Each row shows a different stage, with the first stage at the top. We assume that all of the flip flops have 0 as their output to start with.

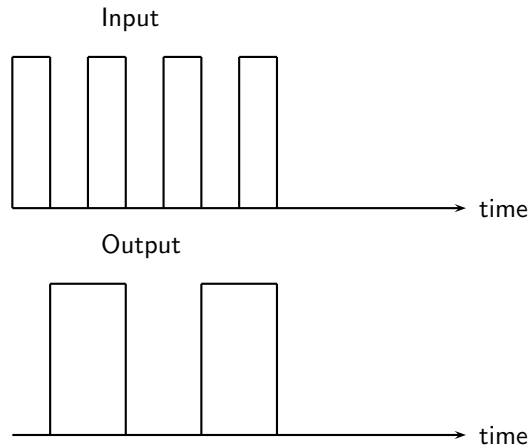


Figure 14.21: The output of a T flip flop, or 'divide by two' circuit when a square wave is connected to the input. The output changes state when the input goes from 1 to 0.

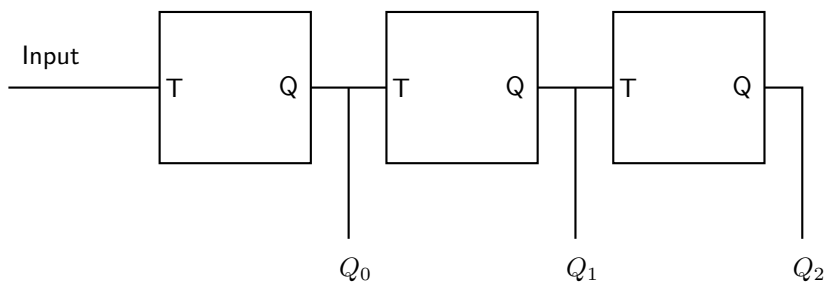


Figure 14.22: Three T flip flops connected together in a chain to make a counter. The input of each flip flop is labelled T, while each output is labelled Q. The pulses are connected to the input on the left. The outputs Q_0 , Q_1 and Q_2 give the three digits of the binary number as the pulses are counted. This is explained in the text and in the next table.

Input	Output 1	Output 2	Output 3	Number of pulse	Number in binary
1	0	0	0	0	000
0	1	0	0	1	001
1	1	0	0	1	001
0	0	1	0	2	010
1	0	1	0	2	010
0	1	1	0	3	011
1	1	1	0	3	011
0	0	0	1	4	100
1	0	0	1	4	100
0	1	0	1	5	101
1	1	0	1	5	101
0	0	1	1	6	110
1	0	1	1	6	110
0	1	1	1	7	111
1	1	1	1	7	111
0	0	0	0	8	1000
1	0	0	0	8	1000
0	1	0	0	9	1101
1	1	0	0	9	1101

The binary numbers in the right hand column count the pulses arriving at the input. You will notice that the output of the first flip flop gives the right most digit of the pulse count (in binary). The output of the second flip flop gives the second digit from the right (the 'twos' digit) of the pulse count. The output of the third flip flop gives the third digit from the right (the 'fours' digit) of the pulse count. As there are only three flip flops, there is nothing to provide the next digit (the 'eights' digit), and so the eighth pulse is recorded as 000, not 1000.

This device is called a **modulo 8** counter because it can count in eight stages from 000 to 111 before it goes back to 000. If you put four flip flops in the counter, it will count in sixteen stages from 0000 to 1111, and it is called a modulo 16 counter because it counts in sixteen stages before going back to 0000.



Definition: Modulo

The modulo of a counter tells you how many stages (or pulses) it receives before going back to 0 as its output. Thus a modulo 8 counter counts in eight stages 000, 001, 010, 011, 100, 101, 110, 111, then returns to 000 again.



If a counter contains n flip flops, it will be a modulo 2^n counter. It will count from 0 to $2^n - 1$.

14.6.3 Storing binary numbers

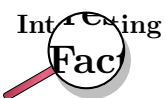
Counting is important. However, it is equally important to be able to remember the numbers. Computers can convert almost anything to a string of 0s and 1s, and therefore to a binary number. Unless this number can be stored in the computer's memory, the computer would be useless.

The memory in the computer contains many parts. Each part is able to store a single 0 or 1. Since 0 and 1 are the two binary digits, we say that each part of the memory stores one **bit**.



Definition: Bit

One bit is a short way of saying one 'binary digit'. It is a single 0 or 1.



If you have eight bits, you can store a binary number from 00000000 to 11111111 (0 to 255 in denary). This gives you enough permutations of 0s and 1s to have one for each letter of the alphabet (in upper and lower case), each digit from 0 to 9, each punctuation mark and each control code used by a computer in storing a document. When you type text into a word processor, each character is stored as a set of eight bits. Each set of eight bits is called a **byte**. Computer memories are graded according to how many bytes they store. There are 1024 bytes in a kilobyte (kB), 1024×1024 bytes in a megabyte (MB), and $1024 \times 1024 \times 1024$ bytes in a gigabyte (GB).

To store a bit we need a circuit which can 'remember' a 0 or a 1. This is called a **bistable** circuit because it has two stable states. It can stay indefinitely either as a 0 or a 1. An example of a bistable circuit is shown in Figure 14.23. It is made from two NOR gates.

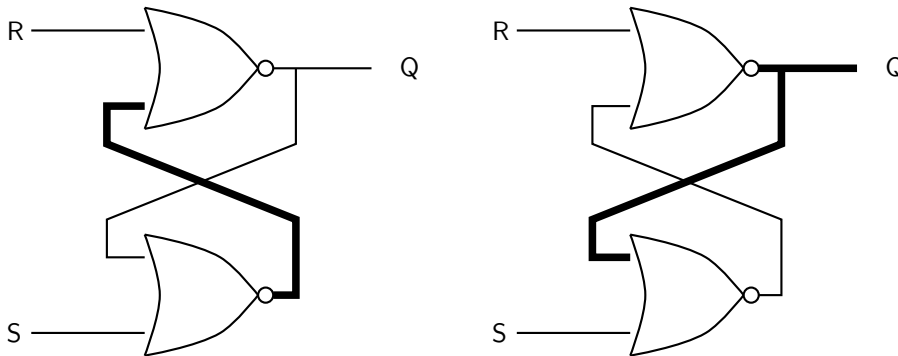


Figure 14.23: A bistable circuit made from two NOR gates. This circuit is able to store one bit of digital information. With the two inputs set to 0, you can see that the output could be (and will remain) either 0 or 1. The circuit on the left shows an output of 0, the circuit on the right shows an output of 1. Wires carrying high logic levels (1) are drawn thicker. The output of the bistable is labelled Q.

To store the 0 or the 1 in the bistable circuit, you set one of the inputs to 1, then put it back to 0 again. If the input labelled 'S' (set) is raised, the output will immediately become 1. This is shown in Figure 14.24.

To store a 0, you raise the 'R' (reset) input to 1. This is shown in Figure 14.25.

Once you have used the S or R inputs to set or reset the bistable circuit, you then bring both inputs back to 0. The bistable 'remembers' the state. Because of the ease with which the circuit can be Reset and Set it is also called an **RS flip flop** circuit.

Computer memory can store millions or billions of bits. If it used our circuit above, it would need millions or billions of NOR gates, each of which is made from several transistors. Computer memory is made of many millions of transistors.

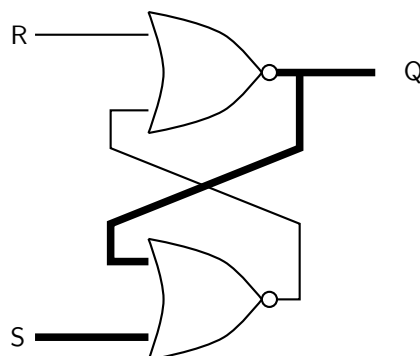


Figure 14.24: The output of a bistable circuit is **set** (made 1) by raising the 'S' input to 1. Wires carrying high logic levels (1) are shown with thicker lines.

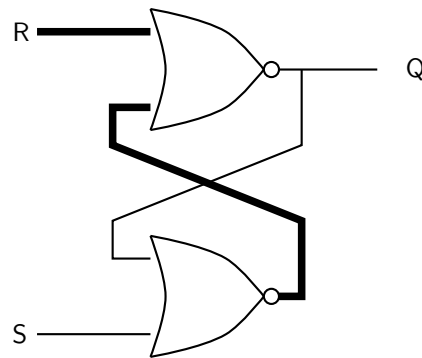


Figure 14.25: The output of a bistable circuit is **reset** (made 0) by raising the 'R' input to 1. Wires carrying high logic levels (1) are shown with thicker lines.



The bistable circuits drawn here don't remember 0s or 1s forever — they lose the information if the power is turned off. The same is true for the RAM (Random Access Memory) used to store working and temporary data in a computer. Some modern circuits contain special memory which can remember its state even if the power is turned off. This is used in FLASH drives, commonly found in USB data sticks and on the memory cards used with digital cameras. These bistable circuits are much more complex.

You can also make T flip flops out of logic gates, however these are more complicated to design.



Exercise: Counting Circuits

1. What is the term *bit* short for?
2. What is 43 in binary?
3. What is 1100101 in denary?
4. What is the highest number a modulo 64 counter can count to? How many T flip flops does it contain?
5. What is the difference between an RS flip flop and a T flip flop?
6. Draw a circuit diagram for a bistable circuit (RS flip flop). Make three extra copies of your diagram. On the first diagram, colour in the wires which will carry high voltage levels (digital 1) if the R input is low, and the S input is high. On the second diagram, colour in the wires which carry high voltage levels if the S input of the first circuit is now made low. On the third diagram, colour in the wires which carry high voltage levels if the R input is now made high. On the final diagram, colour in the wires carrying high voltage levels if the R input is now made low again.
7. Justify the statement: a modern computer contains millions of transistors.



Exercise: End of Chapter Exercises

1. Calculate the reactance of a 3 mH inductor at a frequency of 50 Hz.
 2. Calculate the reactance of a 30 μF capacitor at a frequency of 1 kHz.
 3. Calculate the impedance of a series circuit containing a 5 mH inductor, a 400 μF capacitor and a 2 k Ω resistor at a frequency of 50 kHz.
 4. Calculate the frequency at which the impedance of the circuit in the previous question will be the smallest.
 5. Which component can be used to block low frequencies?
 6. Draw a circuit diagram with a battery, diode and resistor in series. Make sure that the diode is forward biased so that a current will flow through it.
 7. When building a complex electronic circuit which is going to be powered by a battery, it is always a good idea to put a diode in series with the battery. Explain how this will protect the circuit if the user puts the battery in the wrong way round.
 8. Summarize the differences between a bipolar and field effect transistor.
 9. What does an operational amplifier (op-amp) do?
 10. What is the difference between a digital signal and an analogue signal?
 11. What are the advantages of digital signals over analogue signals?
 12. Draw the symbols for the five logic gates, and write down their truth tables.
 13. Draw a circuit diagram with an AND gate. Each input should be connected to the output of a separate NOT gate. By writing truth tables show that this whole circuit behaves as a NOR gate.
 14. Convert the denary number 99 into binary.
 15. Convert the binary number 11100111 into denary.
 16. Explain how three T flip flops can be connected together to make a modulo 8 counter. What is the highest number it can count up to?
 17. Draw the circuit diagram for an RS flip flop (bistable) using two NOR gates.
 18. Show how the circuit you have just drawn can have a stable output of 0 or 1 when both inputs are 0.
 19. Operational (and other) amplifiers, logic gates, and flip flops all contain transistors, and would not work without them. Write a short newspaper article for an intelligent reader who knows nothing about electronics. Explain how important transistors are in modern society.
-

Chapter 15

Electromagnetic Radiation

15.1 Introduction

This chapter will focus on the electromagnetic (EM) radiation. Electromagnetic radiation is a self-propagating wave in space with electric and magnetic components. These components oscillate at right angles to each other and to the direction of propagation, and are in phase with each other. Electromagnetic radiation is classified into types according to the frequency of the wave: these types include, in order of increasing frequency, radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays and gamma rays.

15.2 Particle/wave nature of electromagnetic radiation

If you watch a colony of ants walking up the wall, they look like a thin continuous black line. But as you look closer, you see that the line is made up of thousands of separated black ants.

Light and all other types of electromagnetic radiation seems like a continuous wave at first, but when one performs experiments with light, one can notice that light can have both wave and particle like properties. Just like the individual ants, the light can also be made up of individual bundles of energy, or quanta of light.

Light has both wave-like and particle-like properties (wave–particle duality), but only shows one or the other, depending on the kind of experiment we perform. A wave-type experiment shows the wave nature, and a particle-type experiment shows particle nature. One cannot test the wave and the particle nature at the same time. A particle of light is called a photon.

**Definition: Photon**

A photon is a quantum (energy packet) of light.

The particle nature of light can be demonstrated by the interaction of photons with matter. One way in which light interacts with matter is via the photoelectric effect, which will be studied in detail in Chapter 16.



Exercise: Particle/wave nature of electromagnetic radiation

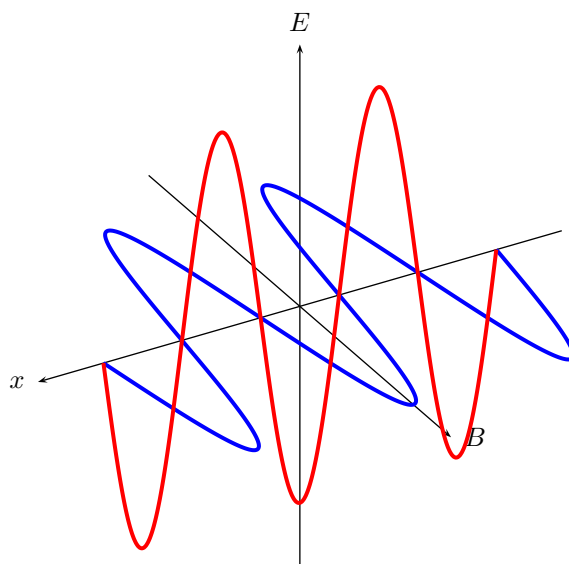
1. Give examples of the behaviour of EM radiation which can best be explained using a wave model.

2. Give examples of the behaviour of EM radiation which can best be explained using a particle model.
-

15.3 The wave nature of electromagnetic radiation

Accelerating charges emit electromagnetic waves. We have seen that a changing electric field generates a magnetic field and a changing magnetic field generates an electric field. This is the principle behind the propagation of electromagnetic waves, because electromagnetic waves, unlike sound waves, do not need a medium to travel through. EM waves propagate when an electric field oscillating in one plane produces a magnetic field oscillating in a plane at right angles to it, which produces an oscillating electric field, and so on. The propagation of electromagnetic waves can be described as *mutual induction*.

These mutually regenerating fields travel through empty space at a constant speed of $3 \times 10^8 \text{ m} \cdot \text{s}^{-1}$, represented by c .



15.4 Electromagnetic spectrum

Observe the things around you, your friend sitting next to you, a large tree across the field. How is it that you are able to see these things? What is it that is leaving your friend's arm and entering your eye so that you can see his arm? It is light. The light originally comes from the sun, or possibly a light bulb or burning fire. In physics, light is given the more technical term electromagnetic radiation, which includes all forms of light, not just the form which you can see with your eyes.

Electromagnetic radiation allows us to observe the world around us. It is this radiation which reflects off of the objects around you and into your eye. The radiation your eye is sensitive to is only a small fraction of the total radiation emitted in the physical universe. All of the different fractions taped together make up the electromagnetic spectrum.



Extension: Dispersion

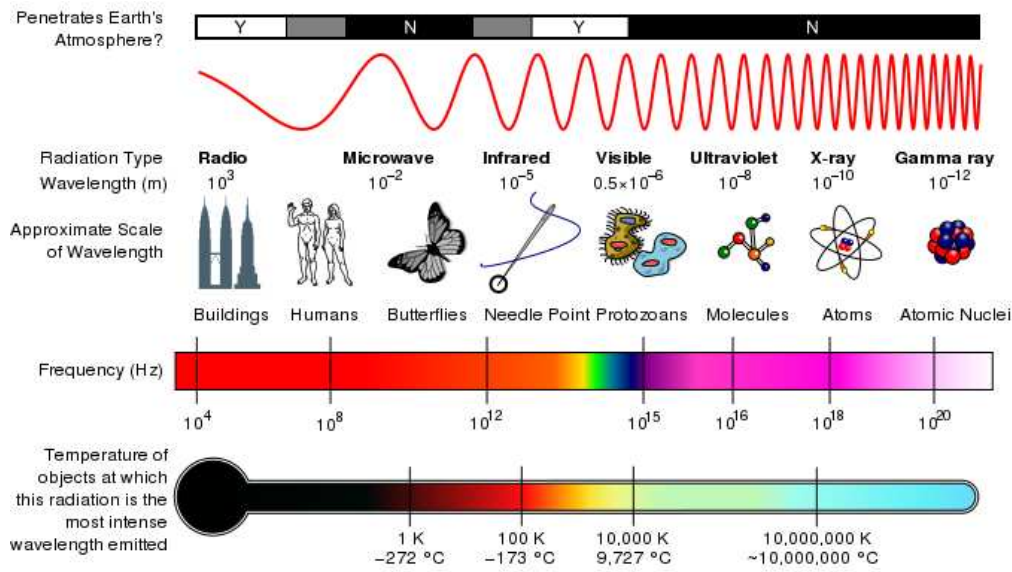


Figure 15.1: The electromagnetic spectrum as a function of frequency. The different types according to wavelength are shown as well as everyday comparisons.

When white light is split into its component colours by a prism, you are looking at a portion of the electromagnetic spectrum.

The wavelength of a particular electromagnetic radiation will depend on how it was created.



Exercise: Wave Nature of EM Radiation

1. List one source of electromagnetic waves. Hint: consider the spectrum diagram and look at the names we give to different wavelengths.
2. Explain how an EM wave propagates, with the aid of a diagram.
3. What is the speed of light? What symbol is used to refer to the speed of light? Does the speed of light change?
4. Do EM waves need a medium to travel through?

The radiation can take on any wavelength, which means that the spectrum is continuous. Physicists broke down this continuous band into sections. Each section is defined by how the radiation is created, not the wavelength of the radiation. But each category is continuous within the min and max wavelength of that category, meaning there are no wavelengths excluded within some range.

The spectrum is in order of wavelength, with the shortest wavelength at one end and the longest wavelength at the other. The spectrum is then broken down into categories as detailed in Table 15.1.

Since an electromagnetic wave is still a wave, the following equation that you learnt in Grade 10 still applies:

$$c = f \cdot \lambda$$

Table 15.1: Electromagnetic spectrum

Category	Range of Wavelengths (nm)	Range of Frequencies (Hz)
gamma rays	<1	$> 3 \times 10^{19}$
X-rays	1-10	3×10^{17} - 3×10^{19}
ultraviolet light	10-400	$7,5 \times 10^{14}$ - 3×10^{17}
visible light	400-700	$4,3 \times 10^{14}$ - $7,5 \times 10^{14}$
infrared	700- 10^5	3×10^{12} - $4,3 \times 10^{19}$
microwave	$10^5 - 10^8$	3×10^9 - 3×10^{12}
radio waves	$> 10^8$	$< 3 \times 10^9$

**Worked Example 82: EM spectrum I**

Question: Calculate the frequency of red light with a wavelength of $4,2 \times 10^{-7}$ m

Answer

We use the formula: $c = f\lambda$ to calculate frequency. The speed of light is a constant 3×10^8 m/s.

$$\begin{aligned} c &= f\lambda \\ 3 \times 10^8 &= f \times 4,2 \times 10^{-7} \\ f &= 7,14 \times 10^{14} \text{ Hz} \end{aligned}$$

**Worked Example 83: EM spectrum II**

Question: Ultraviolet radiation has a wavelength of 200 nm. What is the frequency of the radiation?

Answer

Step 1 : To calculate the frequency we need to identify the wavelength and the velocity of the radiation.

Recall that all radiation travels at the speed of light (c) in vacuum. Since the question does not specify through what type of material the Ultraviolet radiation is traveling, one can assume that it is traveling through a vacuum. We can identify two properties of the radiation - *wavelength* (200 nm) and speed (c).

Step 2 : We can use the equation $c = f\lambda$ to find the frequency since the wavelength is given.

$$\begin{aligned} c &= f\lambda \\ 3 \times 10^8 &= f \times 200 \times 10^{-9} \\ f &= 1,5 \times 10^{15} \text{ Hz} \end{aligned}$$

Examples of some uses of electromagnetic waves are shown in Table 15.2.

In theory the spectrum is infinite, although realistically we can only observe wavelengths from a few hundred kilometers to those of gamma rays due to experimental limitations.

Table 15.2: Uses of EM waves

Category	Uses
gamma rays	used to kill the bacteria in marsh-mallows and to sterilise medical equipment
X-rays	used to image bone structures
ultraviolet light	bees can see into the ultraviolet because flowers stand out more clearly at this frequency
visible light	used by humans to observe the world
infrared	night vision, heat sensors, laser metal cutting
microwave	microwave ovens, radar
radio waves	radio, television broadcasts

Humans experience electromagnetic waves differently depending on their wavelength. Our eyes are sensitive to visible light while our skin is sensitive to infrared, and many wavelengths we do not detect at all.



Exercise: EM Radiation

1. Arrange the following types of EM radiation in order of increasing frequency: infrared, X-rays, ultraviolet, visible, gamma.
2. Calculate the frequency of an EM wave with a wavelength of 400 nm.
3. Give an example of the use of each type of EM radiation, i.e. gamma rays, X-rays, ultraviolet light, visible light, infrared, microwave and radio and TV waves.

15.5 The particle nature of electromagnetic radiation

When we talk of electromagnetic radiation as a particle, we refer to photons, which are packets of energy. The energy of the photon is related to the wavelength of electromagnetic radiation according to: h (called Planck's constant).

Definition: Planck's constant

Planck's constant is a physical constant named after Max Planck.

$$h = 6,626 \times 10^{-34} \text{ J} \cdot \text{s}$$

The energy of a photon can be calculated using the formula: $E = hf$ or $E = h\frac{c}{\lambda}$. Where E is the energy of the photon in joules (J), h is planck's constant, c is the speed of light, f is the frequency in hertz (Hz) and λ is the wavelength in metres (m).

Worked Example 84: Calculating the energy of a photon I**Question:** Calculate the energy of a photon with a frequency of 3×10^{18} Hz**Answer**We use the formula: $E = hf$

$$\begin{aligned}
 E &= hf \\
 &= 6,6 \times 10^{-34} \times 3 \times 10^{18} \\
 &= 2 \times 10^{-15} \text{ J}
 \end{aligned}$$

**Worked Example 85: Calculating the energy of a photon II****Question:** What is the energy of an ultraviolet photon with a wavelength of 200 nm?**Answer****Step 1 : Determine what is required and how to approach the problem.**

We are required to calculate the energy associated with a photon of ultraviolet light with a wavelength of 200 nm.

We can use:

$$E = h \frac{c}{\lambda}$$

Step 2 : Solve the problem

$$\begin{aligned}
 E &= h \frac{c}{\lambda} \\
 &= (6,626 \times 10^{-34}) \frac{3 \times 10^8}{200 \times 10^{-9}} \\
 &= 9,939 \times 10^{-10} \text{ J}
 \end{aligned}$$

15.5.1 Exercise - particle nature of EM waves

1. How is the energy of a photon related to its frequency and wavelength?
2. Calculate the energy of a photon of EM radiation with a frequency of 10^{12} Hz.
3. Determine the energy of a photon of EM radiation with a wavelength of 600 nm.

15.6 Penetrating ability of electromagnetic radiation

Different kinds of electromagnetic radiation have different penetrabilities. For example, if we take the human body as the object. Infrared light is emitted by the human body. Visible light is reflected off the surface of the human body, ultra-violet light (from sunlight) damages the skin, but X-rays are able to penetrate the skin and bone and allow for pictures of the inside of the human body to be taken.

If we compare the energy of visible light to the energy of X-rays, we find that X-rays have a much higher energy. Usually, kinds of electromagnetic radiation with higher energy have higher penetrabilities than those with low energies.

Certain kinds of electromagnetic radiation such as ultra-violet radiation, X-rays and gamma rays are very dangerous. Radiation such as these are called ionising radiation. Ionising radiation transfers energy as it passes through matter, breaking molecular bonds and creating ions.

Excessive exposure to radiation, including sunlight, X-rays and all nuclear radiation, can cause destruction of biological tissue.

15.6.1 Ultraviolet(UV) radiation and the skin

UVA and UVB are different ranges of frequencies for ultraviolet (UV) light. UVA and UVB can damage collagen fibres which results in the speeding up skin aging. In general, UVA is the least harmful, but it can contribute to the aging of skin, DNA damage and possibly skin cancer. It penetrates deeply and does not cause sunburn. Because it does not cause reddening of the skin (erythema) it cannot be measured in the SPF testing. There is no good clinical measurement of the blocking of UVA radiation, but it is important that sunscreen block both UVA and UVB.

UVB light can cause skin cancer. The radiation excites DNA molecules in skin cells, resulting in possible mutations, which can cause cancer. This cancer connection is one reason for concern about ozone depletion and the ozone hole.

As a defense against UV radiation, the body tans when exposed to moderate (depending on skin type) levels of radiation by releasing the brown pigment melanin. This helps to block UV penetration and prevent damage to the vulnerable skin tissues deeper down. Suntan lotion, often referred to as sunblock or sunscreen, partly blocks UV and is widely available. Most of these products contain an SPF rating that describes the amount of protection given. This protection, however, applies only to UVB rays responsible for sunburn and not to UVA rays that penetrate more deeply into the skin and may also be responsible for causing cancer and wrinkles. Some sunscreen lotion now includes compounds such as titanium dioxide which helps protect against UVA rays. Other UVA blocking compounds found in sunscreen include zinc oxide and avobenzene.



Extension: What makes a good sunscreen?

- UVB protection: Padimate O, Homosalate, Octisalate (octyl salicylate), Octinoxate (octyl methoxycinnamate)
- UVA protection: Avobenzene
- UVA/UVB protection: Octocrylene, titanium dioxide, zinc oxide, Mexoryl (ecamsule)

Another means to block UV is by wearing sun protective clothing. This is clothing that has a UPF rating that describes the protection given against both UVA and UVB.

15.6.2 Ultraviolet radiation and the eyes

High intensities of UVB light are hazardous to the eyes, and exposure can cause welder's flash (photo keratitis or arc eye) and may lead to cataracts, pterygium and pinguecula formation.

Protective eyewear is beneficial to those who are working with or those who might be exposed to ultraviolet radiation, particularly short wave UV. Given that light may reach the eye from the sides, full coverage eye protection is usually warranted if there is an increased risk of exposure, as in high altitude mountaineering. Mountaineers are exposed to higher than ordinary levels of UV radiation, both because there is less atmospheric filtering and because of reflection from snow and ice.

Ordinary, untreated eyeglasses give some protection. Most plastic lenses give more protection than glass lenses. Some plastic lens materials, such as polycarbonate, block most UV. There are protective treatments available for eyeglass lenses that need it which will give better protection. But even a treatment that completely blocks UV will not protect the eye from light that arrives around the lens. To convince yourself of the potential dangers of stray UV light, cover your lenses with something opaque, like aluminum foil, stand next to a bright light, and consider how much light you see, despite the complete blockage of the lenses. Most contact lenses help to protect the retina by absorbing UV radiation.

15.6.3 X-rays

While x-rays are used significantly in medicine, prolonged exposure to X-rays can lead to cell damage and cancer.

For example, a mammogram is an x-ray of the human breast to detect breast cancer, but if a woman starts having regular mammograms when she is too young, her chances of getting breast cancer increases.

15.6.4 Gamma-rays

Due to the high energy of gamma-rays, they are able to cause serious damage when absorbed by living cells.

Gamma-rays are not stopped by the skin and can induce DNA alteration by interfering with the genetic material of the cell. DNA double-strand breaks are generally accepted to be the most biologically significant lesion by which ionising radiation causes cancer and hereditary disease.

A study done on Russian nuclear workers exposed to external whole-body gamma-radiation at high cumulative doses shows a link between radiation exposure and death from leukaemia, lung, liver, skeletal and other solid cancers.



Extension: Cellphones and electromagnetic radiation

Cellphone radiation and health concerns have been raised, especially following the enormous increase in the use of wireless mobile telephony throughout the world. This is because mobile phones use electromagnetic waves in the microwave range. These concerns have induced a large body of research. Concerns about effects on health have also been raised regarding other digital wireless systems, such as data communication networks. In 2009 the World Health Organisation announced that they have found a link between brain cancer and cellphones.

Cellphone users are recommended to minimise radiation, by for example:

1. Use hands-free to decrease the radiation to the head.
2. Keep the mobile phone away from the body.
3. Do not telephone in a car without an external antenna.

15.6.5 Exercise - Penetrating ability of EM radiation

1. Indicate the penetrating ability of the different kinds of EM radiation and relate it to energy of the radiation.
2. Describe the dangers of gamma rays, X-rays and the damaging effect of ultra-violet radiation on skin

15.7 Summary

1. Electromagnetic radiation has both a wave and a particle nature.
2. Electromagnetic waves travel at a speed of $3 \times 10^8 \text{ m} \cdot \text{s}^{-1}$ in a vacuum.
3. The Electromagnetic spectrum consists of the following types of radiation: radio waves, microwaves, infrared, visible, ultraviolet, X-rays, gamma-rays.
4. Gamma-rays have the most energy and are the most penetrating, while radio waves have the lowest energy and are the least penetrating.

15.8 End of chapter exercise

1. What is the energy of a photon of EM radiation with a frequency of $3 \times 10^8 \text{ Hz}$?
2. What is the energy of a photon of light with a wavelength of 660 nm?
3. List the main types of electromagnetic radiation in order of increasing wavelength.
4. List the main uses of:
 - (a) radio waves
 - (b) infrared
 - (c) gamma rays
 - (d) X-rays
5. Explain why we need to protect ourselves from ultraviolet radiation from the Sun.
6. List some advantages and disadvantages of using X-rays.
7. What precautions should we take when using cell phones?
8. Write a short essay on a type of electromagnetic waves. You should look at uses, advantages and disadvantages of your chosen radiation.
9. Explain why some types of electromagnetic radiation are more penetrating than others.

Chapter 16

Optical Phenomena and Properties of Matter - Grade 12

16.1 Introduction

For centuries physicists argued about whether light was a particle or a wave. It was assumed that light could only be one or the other, but not both.

In earlier chapters on waves and optics, you studied how light or other electromagnetic radiation propagates like a wave. The wave nature of light was demonstrated by the propagation of light in examples such as diffraction, interference, and polarisation of light.

You also saw in Electromagnetic Radiation how light sometimes behaves as a particle. This chapter looks at evidence supporting the *particle model of light*. The idea that light can have both wave and particle properties was one of the most important discoveries of the twentieth century.

16.2 The transmission and scattering of light

16.2.1 Energy levels of an electron

We have seen that the electrons in an atom have different energy levels. When the electron receives enough energy, it can jump up to a higher energy level. This is called '*exciting*' the electron. When the electron in a high energy level sheds some energy, it drops to a lower energy level.

We have also seen that the energy associated with light at a specific wavelength is given by:

$$E = \frac{hc}{\lambda}.$$

In the particle model of light, this means that each packet of light (photon) at a wavelength λ has energy:

$$E = \frac{hc}{\lambda}.$$

For the electron to receive energy, it *absorbs* a photon and gets its energy. When an electron loses energy to drop to a lower level, it *emits* (gives off) a photon with that energy.

16.2.2 Interaction of light with metals

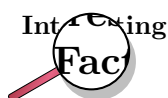
When light encounters or passes through a material, the photons of the light interact with the atoms or molecules of the material. Depending on the strength of the interactions and how often they happen, the light will pass through the material or be scattered in some other direction.

Each wavelength of light relates to a particular energy, and the closer that energy is to the energy difference between two of the levels of the atom, the likelier the photon is to interact with the atom.

When visible or ultraviolet (UV) radiation shines on a metal, the photons are *absorbed* by the electrons in the metal. The electrons are then excited up to a higher energy level. When an electron returns to a lower energy level, another photon is emitted. This is how light appears to be **reflected** off a metal surface.

In previous chapters, you have studied geometrical optics, which tells us what happens to rays of light when they are reflected off a surface or refracted through a lens. That tells us what happens to light rays, made up of many photons, on a large scale. If you look at a smaller level, i.e. on a microscopic scale, then reflection and refraction happen by all the photons interacting with the atoms of the lens or mirror. The photons get absorbed and re-emitted many times before emerging as the final rays of light that we see.

Scattering of light is responsible for many effects in everyday life. We see that certain materials are red or blue, for example, since they contain materials that have energy level differences that correspond to the energies of the photons that make up red or blue light. These materials then reflect the red or blue light and absorb the other wavelengths in the visible spectrum. White objects reflect photons of all wavelengths in the visual spectrum, while black objects absorb these photons.



Because a truly black object absorbs all the visual wavelengths of light, and does not re-emit photons at visual wavelengths, we can say that 'black' is not a colour itself, but rather a *lack of* colour! Also, since black objects absorb visual light, they heat up more than objects of other colours which reflect light at certain wavelengths.

Activity :: Investigation : Reflection and absorption

Aim:

Investigate the interaction of light with differently coloured metal objects

Apparatus:

Find some differently coloured metal objects (at least 5) which will not be damaged if left in the sun for 15 minutes. Make sure to include at least one white item and one black item.

Method:

At the start of your lesson set out your objects in direct sunlight. Leave them there for around 15 minutes.

Alternatively, if it is a sunny day, you can use your teachers' cars for this experiment - as long as there are some cars of different colours and they have been standing in direct sunlight for the same length of time!

After 15 minutes is up, touch each of the items/cars (be careful not to burn yourself!) and compare their temperatures (rate them from 1 to 5 with 1 being cold and 5 being very hot) in a table such as the example table below:

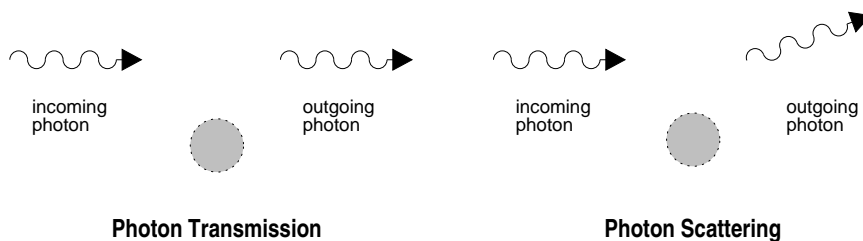


Figure 16.1: Diagrams of photon transmission (left) and scattering (right).

Object	Colour	Temperature rating
e.g. car 1	e.g. red	e.g. 3

Questions:

1. Which object was the hottest and what was it made of?
2. Which object was the coolest and what was it made of?
3. How did the temperatures of the black and white objects compare to each other? (which one was hotter and which was cooler?)
4. Try to explain the reasons for the different temperatures of the objects with respect to their colours and the materials of which they are made.

Metals generally reflect most wavelengths of visible light, but they will reflect the light in a certain direction, given by the laws of reflection in geometrical optics. This is different to most materials, like wood or fabric, which reflect light in all directions. Metals have this property since they have electrons that are not bound to atoms and can move freely through the metal. This is unlike most other materials that have their electrons bound closely to the atoms. These free electrons in metals can then absorb and reflect photons of a wide range of energies.

Ultraviolet light (which has shorter wavelengths than visible light) will pass through some substances, such as many plastics, because they do not have the right energy levels to absorb it and re-emit it. *X-rays* (also short wavelengths) will also pass through most materials, since the energies of X-rays correspond to the energy levels of atomic nuclei. Such nuclei are much smaller than atoms, so it is much less likely for an X-ray to hit a nucleus instead of the whole atom.

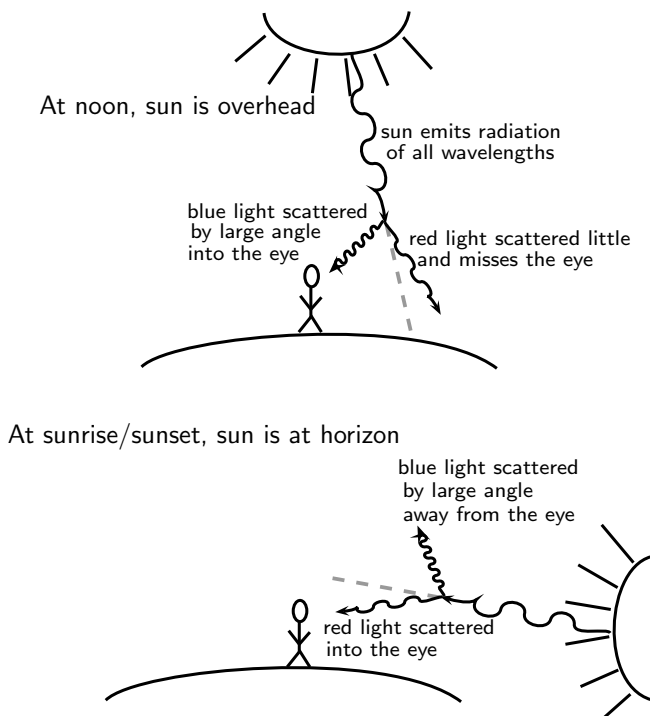
Most materials will absorb *infrared radiation* (longer wavelengths than visible light), since the energies of that radiation often correspond to rotational or vibrational energy levels of molecules.

16.2.3 Why is the sky blue?

The sun emits light at many different wavelengths, including all of the visible wavelengths. Light which is made up of all the visible wavelengths appears white. So what causes the sky to look blue?

The atmosphere consists of molecules of different gases as well as tiny dust grains. Light from the sun scatters off the molecules in the air (called **Rayleigh scattering**). The chance that the light will scatter off the gas molecules is *higher* for *shorter* wavelengths. The *short wavelength blue light* is therefore scattered more than the other colours. At noon, when the light from the sun is coming straight down (see the picture), the scattered blue light reaches your eyes from all directions and so the sky appears blue. The other wavelengths do not get scattered much and therefore miss your line of sight and are not seen. At sunrise or sunset, the direction of the light coming from the sun is now straight towards your eyes (see the picture). Therefore the scattered blue light can't be seen because it is scattered out of your line of sight. The redder

colours (oranges and reds) can now be seen because they are not scattered as much and still fall in your line of sight.



Exercise: Transmission and scattering of light

1. Explain how visible light is reflected from metals.
2. Explain why the sky is blue.

16.3 The photoelectric effect

Around the turn of the twentieth century, it was observed by a number of physicists (including Hertz, Thomson and Von Lenard) that when light was shone on a metal, electrons were emitted by the metal. This is called the photoelectric effect. (*photo-* for light, *electric-* for the electron.)



Definition: The photoelectric effect

The photoelectric effect is the process whereby an electron is emitted by a metal when light shines on it.

At that time, light was thought to be purely a wave. Therefore, physicists thought that if a more intense (i.e. brighter) light was shone on a metal, then the electrons would be knocked out with greater kinetic energies than if a faint light was shone on them. However, Von Lenard observed that this did not happen at all. The *intensity* of the light made *no difference* to the kinetic energy of the emitted electrons! Also, it was observed that the electrons were emitted immediately when light was shone on the metal - there was no time delay.

Einstein solved this problem by proposing that light is made up of packets of energy called *quanta* (now called photons) which interacted with the electrons in the metal like particles instead of waves. Each incident photon would transfer all its energy to one electron in the metal. For a specific colour of light (i.e. a certain wavelength or frequency), the energy of the photons is given by $E = hf = hc/\lambda$, where h is Planck's constant. The energy needed to knock an electron out of the metal is called the **work function** (symbol ϕ) of the metal. Therefore, the amount of energy left over as the kinetic energy (E_k) of the emitted electron would be the difference between the incoming photon's energy and the energy needed to knock out the electron (work function of the metal):

$$E_k = hf - \phi$$

Increasing the intensity of the light (i.e. making it brighter) did not change the wavelength of the light and therefore the electrons would be emitted with *the same* kinetic energy as before! This solved the paradox and showed that light has *both* a **wave nature** and a **particle nature**. Einstein won the Nobel prize for this quantum theory and his explanation of the photoelectric effect.

Increasing the intensity of the light actually means increasing the *number* of incident photons. Therefore, since each photon only gives energy to one electron, more incident photons means *more* electrons would be knocked out of the metal, but their kinetic energies would be *the same* as before.

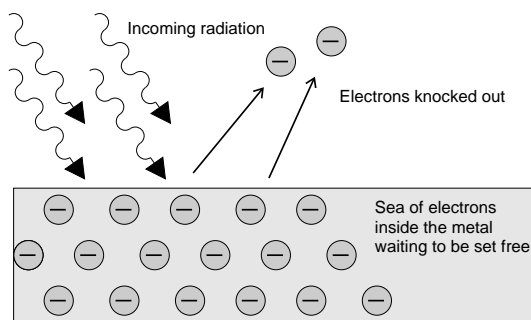
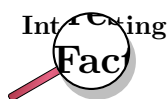


Figure 16.2: The photoelectric effect: Incoming photons on the left hit the electrons inside the metal surface. The electrons absorb the energy from the photons, and are ejected from the metal surface.



The photoelectric effect was first observed in the experiments of Heinrich Hertz in 1887. In 1899 J.J. Thomson proved that it was electrons that were emitted. The photoelectric effect was theoretically explained by Albert Einstein in 1905.

The discovery and understanding of the photoelectric effect was one of the major breakthroughs in science in the twentieth century as it provided concrete evidence of the particle nature of light. It overturned previously held views that light was composed purely of a continuous transverse wave. On the one hand, the wave nature is a good description of phenomena such as diffraction and interference for light, and on the other hand, the photoelectric effect demonstrates the particle nature of light. This is now known as the 'dual-nature' of light. (dual means two)

While solving problems we need to decide for ourselves whether we should consider the wave property or the particle property of light. For example, when dealing with interference and diffraction, light should be treated as a wave, whereas when dealing with photoelectric effect we consider the particle nature.

16.3.1 Applications of the photoelectric effect

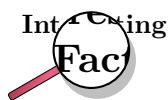
We have learnt that a metal contains electrons that are free to move between the valence and conduction bands. When a photon strikes the surface of a metal, it gives *all* its energy to one electron in the metal.

- If the photon energy is *equal* to the energy between two energy levels then the electron is excited to the higher energy level.
- If the photon energy is *greater* than or *equal* to the work function (energy needed to escape from the metal), then the electron is emitted from the surface of the metal (the photoelectric effect).

The work function is different for different elements. The smaller the work function, the easier it is for electrons to be emitted from the metal. Metals with low work functions make good conductors. This is because the electrons are attached less strongly to their surroundings and can move more easily through these materials. This reduces the resistance of the material to the flow of current i.e. it conducts well. Table 16.1 shows the work functions for a range of elements.

Element	Work Function (J)
Aluminium	$6,9 \times 10^{-19}$
Beryllium	$8,0 \times 10^{-19}$
Calcium	$4,6 \times 10^{-19}$
Copper	$7,5 \times 10^{-19}$
Gold	$8,2 \times 10^{-19}$
Lead	$6,9 \times 10^{-19}$
Silicon	$1,8 \times 10^{-19}$
Silver	$6,9 \times 10^{-19}$
Sodium	$3,7 \times 10^{-19}$

Table 16.1: Work functions of selected elements determined from the photoelectric effect. (From the *Handbook of Chemistry and Physics*.)



The **electron volt** (eV) is the kinetic energy gained by an electron passing through a potential difference of one volt (1 V). A **volt** is not a measure of energy, but the **electron volt** is a unit of energy. When you connect a 1.5 V battery to a circuit, you can give 1.5 eV of energy to every electron.



Worked Example 86: The photoelectric effect - I

Question: Ultraviolet radiation with a wavelength of 250 nm is incident on a silver foil (work function $\phi = 6,9 \times 10^{-19}$). What is the maximum kinetic energy of the emitted electrons?

Answer

Step 1 : Determine what is required and how to approach the problem

We need to determine the maximum kinetic energy of an electron ejected from a silver foil by ultraviolet radiation.

The photoelectric effect tells us that:

$$E_k = E_{\text{photon}} - \phi$$

$$E_k = h \frac{c}{\lambda} - \phi$$

We also have:

Work function of silver: $\phi = 6,9 \times 10^{-19} \text{ J}$

UV radiation wavelength = $250 \text{ nm} = 250 \times 10^{-9} \text{ m}$

Planck's constant: $h = 6,63 \times 10^{-34} \text{ m}^2\text{kgs}^{-1}$

speed of light: $c = 3 \times 10^8 \text{ ms}^{-1}$

Step 2 : Solve the problem

$$E_k = \frac{hc}{\lambda} - \phi$$

$$= [6,63 \times 10^{-34} \times \frac{3 \times 10^8}{250 \times 10^{-9}}] - 6,9 \times 10^{-19}$$

$$= 1,06 \times 10^{-19} \text{ J}$$

The maximum kinetic energy of the emitted electron will be $1,06 \times 10^{-19} \text{ J}$.



Worked Example 87: The photoelectric effect - II

Question: If we were to shine the same ultraviolet radiation ($f = 1,2 \times 10^{15} \text{ Hz}$), on a gold foil (work function = $8,2 \times 10^{-19} \text{ J}$), would any electrons be emitted from the surface of the gold foil?

Answer

For the electrons to be emitted from the surface, the energy of each photon needs to be *greater* than the work function of the material.

Step 1 : Calculate the energy of the incident photons

$$E_{\text{photon}} = hf$$

$$= 6,63 \times 10^{-34} \times 1,2 \times 10^{15}$$

$$= 7,96 \times 10^{-19} \text{ J}$$

Therefore each photon of ultraviolet light has an energy of $7,96 \times 10^{-19} \text{ J}$.

Step 2 : Write down the work function for gold.

$$\phi_{\text{gold}} = 8,2 \times 10^{-19} \text{ J}$$

Step 3 : Is the energy of the photons greater or smaller than the work function?

$$7,96 \times 10^{-19} \text{ J} < 8,2 \times 10^{-19} \text{ J}$$

$$E_{\text{photons}} < \phi_{\text{gold}}$$

The energy of each photon is less than the work function of gold, therefore,

the photons do not have enough energy to knock electrons out of the gold.
No electrons would be emitted from the gold foil.



Extension: Units of energy

When dealing with calculations at a small scale (like at the level of electrons) it is more convenient to use different units for energy rather than the joule (J). We define a unit called the electron-volt (eV) as the kinetic energy gained by an electron passing through a potential difference of one volt.

$$E = q \times V$$

where q is the charge of the electron and V is the potential difference applied. The charge of 1 electron is $1,6 \times 10^{-19}$ C, so 1 eV is calculated to be:

$$1 \text{ eV} = (1,6 \times 10^{-19} \text{ C} \times 1 \text{ V}) = 1,6 \times 10^{-19} \text{ J}$$

You can see that $1,6 \times 10^{-19}$ J is a very small amount of energy and so using electron-volts (eV) at this level is easier.

Hence, $1 \text{ eV} = 1,6 \times 10^{-19} \text{ J}$ which means that $1 \text{ J} = 6,241 \times 10^{18} \text{ eV}$

16.3.2 Real-life applications

Solar Cells

The photo-electric effect may seem like a very easy way to produce electricity from the sun. This is why people choose to make solar panels out of materials like silicon, to generate electricity. In real-life however, the amount of electricity generated is less than expected. This is because not every photon knocks out an electron. Other processes such as reflection or scattering also happen. This means that only a fraction $\approx 10\%$ (depends on the material) of the photons produce photoelectrons. This drop in efficiency results in a lower current. Much work is being done in industry to improve this efficiency so that the panels can generate as high a current as possible, and create as much electricity as possible from the sun. But even these smaller electrical currents are useful in applications like solar-powered calculators.



Exercise: The photoelectric effect

1. Describe the photoelectric effect.
2. List two reasons why the observation of the photoelectric effect was significant.
3. Refer to Table 16.1: If I shine ultraviolet light with a wavelength of 288 nm onto some aluminium foil, what would be the kinetic energy of the emitted electrons?
4. I shine a light of an unknown wavelength onto some silver foil. The light has only enough energy to eject electrons from the silver foil but not enough to give them kinetic energy. (Refer to Table 16.1 when answering the questions below:)
 - (a) If I shine the same light onto some copper foil, would electrons be ejected?
 - (b) If I shine the same light onto some silicon, would electrons be ejected?

- (c) If I increase the intensity of the light shining on the silver foil, what happens?
- (d) If I increase the frequency of the light shining on the silver foil, what happens?

16.4 Emission and absorption spectra

16.4.1 Emission Spectra

You have learnt previously about the structure of an atom. The electrons surrounding the atomic nucleus are arranged in a series of levels of increasing energy. Each element has its own distinct set of energy levels. This arrangement of energy levels serves as the atom's unique fingerprint.

In the early 1900s, scientists found that a liquid or solid heated to high temperatures would give off a broad range of colours of light. However, a gas heated to similar temperatures would emit light only at certain specific colours (wavelengths). The reason for this observation was not understood at the time.

Scientists studied this effect using a discharge tube.

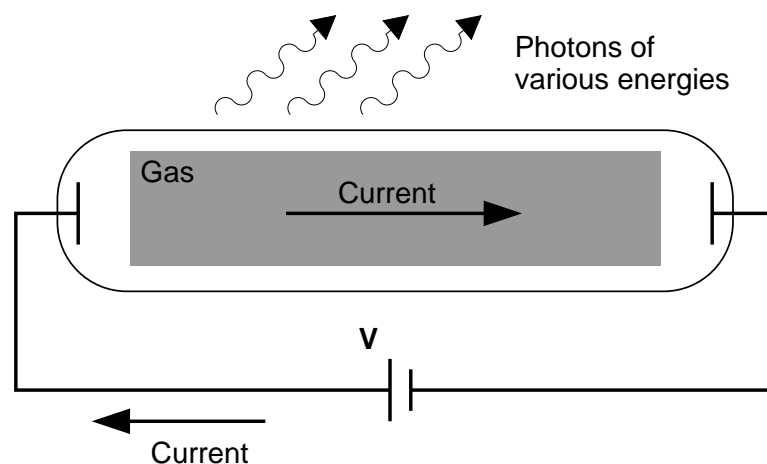


Figure 16.3: Diagram of a discharge tube. The tube is filled with a gas. When a high enough voltage is applied across the tube, the gas ionises and acts like a conductor, allowing a current to flow through the circuit. The current excites the atoms of the ionised gas. When the atoms fall back to their ground state, they emit photons to carry off the excess energy.

A discharge tube (Figure 16.3) is a glass gas-filled tube with a metal plate at both ends. If a large enough voltage difference is applied between the two metal plates, the gas atoms inside the tube will absorb enough energy to make some of their electrons come off i.e. the gas atoms are ionised. These electrons start moving through the gas and create a current, which raises some electrons in other atoms to higher energy levels. Then as the electrons in the atoms fall back down, they emit electromagnetic radiation (light). The amount of light emitted at different wavelengths, called the **emission spectrum**, is shown for a discharge tube filled with hydrogen gas in figure 16.4 below. Only certain wavelengths (i.e. colours) of light are seen as shown by the thick black lines in the picture.

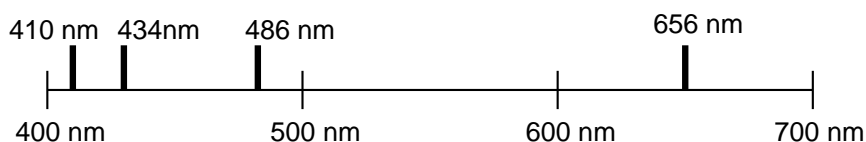


Figure 16.4: Diagram of the emission spectrum of hydrogen in the visible spectrum. Four lines are visible, and are labeled with their wavelengths. The three lines in the 400–500 nm range are in the blue part of the spectrum, while the higher line (656 nm) is in the red/orange part.

Eventually, scientists realized that these lines come from photons of a specific energy, emitted by electrons making transitions between specific energy levels of the atom. Figure ?? shows an example of this happening. When an electron in an atom falls from a higher energy level to a lower energy level, it emits a photon to carry off the extra energy. This photon's energy is equal to the energy difference between the two energy levels. As we previously discussed, the frequency of a photon is related to its energy through the equation $E = hf$. Since a specific photon frequency (or wavelength) gives us a specific colour, we can see how each coloured line is associated with a specific transition.

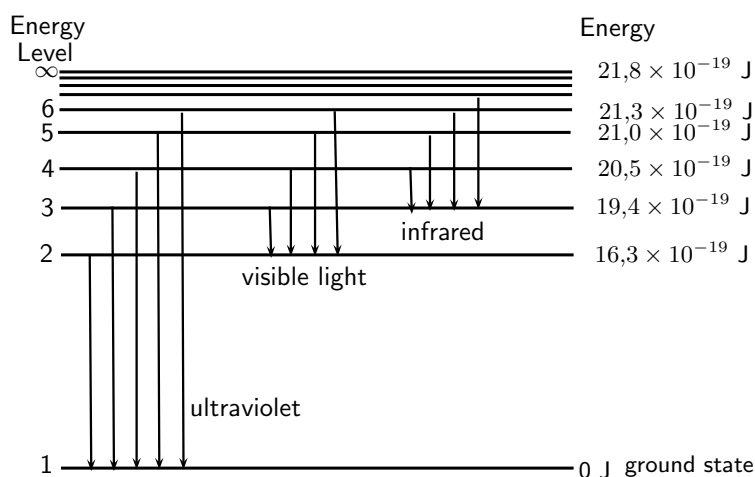


Figure 16.5: In this diagram are shown some of the electron energy levels for the hydrogen atom. The arrows show the electron transitions from higher energy levels to lower energy levels. The energies of the emitted photons are the same as the energy difference between two energy levels. You can think of absorption as the opposite process. The arrows would point upwards and the electrons would jump up to higher levels when they absorb a photon of the right energy.

Visible light is not the only kind of electromagnetic radiation emitted. More energetic or less energetic transitions can produce ultraviolet or infrared radiation. However, because each atom has its own distinct set of energy levels (its fingerprint!), each atom has its own distinct emission spectrum.

16.4.2 Absorption spectra

As you know, atoms do not only emit photons; they also absorb photons. If a photon hits an atom and the energy of the photon is the same as the gap between two electron energy levels in the atom, then the electron can absorb the photon and jump up to the higher energy level. If the atom has no energy level differences that equal the incoming photon's energy, it cannot absorb the photon, and can only scatter it.

Using this effect, if we have a source of photons of various energies we can obtain the **absorption spectra** for different materials. To get an absorption spectrum, just shine white light on a sample of the material that you are interested in. White light is made up of all the different wavelengths

of visible light put together. In the absorption spectrum, the energy levels corresponding to the absorbed photons show up as black lines because the photons of these wavelengths have been absorbed and don't show up. Because of this, the absorption spectrum is the exact *inverse* of the emission spectrum. Look at the two figures below. In figure 16.6 you can see the emission lines of hydrogen. Figure 16.7 shows the absorption spectrum. It is the exact opposite of the emission spectrum! Both emission and absorption techniques can be used to get the same information about the energy levels of an atom.

Figure 16.6: *Emission* spectrum of Hydrogen.Figure 16.7: *Absorption* spectrum of Hydrogen.**Worked Example 88: Absorption**

Question: I have an unknown gas in a glass container. I shine a bright white light through one side of the container and measure the spectrum of transmitted light. I notice that there is a black line (*absorption* line) in the middle of the visible red band at 642 nm. I have a hunch that the gas might be hydrogen. If I am correct, between which 2 energy levels does this transition occur? (Hint: look at figure 16.5 and the transitions which are in the visible part of the spectrum.)

Answer**Step 1 : What is given and what needs to be done?**

We have an absorption line at 642 nm. This means that the substance in the glass container absorbed photons with a wavelength of 642 nm. We need to calculate which 2 energy levels of hydrogen this transition would correspond to. Therefore we need to know what energy the absorbed photons had.

Step 2 : Calculate the energy of the absorbed photons

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{642 \times 10^{-9}} \\ &= 3,1 \times 10^{-19} \text{ J} \end{aligned}$$

The absorbed photons had energy of $3,1 \times 10^{-19}$

Step 3 : Find the energy of the transitions resulting in radiation at visible wavelengths

Figure 16.5 shows various energy level transitions. The transitions related to visible wavelengths are marked as the transitions beginning or ending on Energy Level 2. Let's find the energy of those transitions and compare with the energy of the absorbed photons we've just calculated.

Energy of transition (absorption) from Energy Level 2 to Energy Level 3:

$$\begin{aligned} E_{2,3} &= E_2 - E_3 \\ &= 16,3 \times 10^{-19} \text{ J} - 19,4 \times 10^{-19} \text{ J} \\ &= -3,1 \times 10^{-19} \text{ J} \end{aligned}$$

Therefore the energy of the photon that an electron must absorb to jump

from Energy Level 2 to Energy Level 3 is $3,1 \times 10^{-19}$ J. (NOTE: The minus sign means that *absorption* is occurring.)

This is the same energy as the photons which were absorbed by the gas in the container! Therefore, since the transitions of all elements are unique, we can say that the gas in the container is hydrogen. The transition is absorption of a photon between Energy Level 2 and Energy Level 3.

16.4.3 Colours and energies of electromagnetic radiation

We saw in the explanation for why the sky is blue that different *wavelengths* or *frequencies* of light correspond to different *colours* of light. The table below gives the wavelengths and colours of light in the visible spectrum:

Colour	Wavelength range (nm)
violet	390 - 455
blue	455 - 492
green	492 - 577
yellow	577 - 597
orange	597 - 622
red	622 - 780

Table 16.2: Colours and wavelengths of light in the visible spectrum.

We also know that the energy of a photon of light can be found from:

$$E = hf = \frac{hc}{\lambda}$$

Therefore if we know the frequency or wavelength of light, we can calculate the photon's energy and vice versa.

Activity :: Investigation : Frequency, wavelength and energy relation

Refer to table 16.2: Copy the table into your workbook and add two additional columns.

1. In the first new column write down the lower and upper frequencies for each colour of light.
2. In the second column write down the energy range (in Joules) for each colour of light.

Questions

1. Which colour of visible light has the highest energy photons?
 2. Which colour of visible light has the lowest energy photons?
-



Worked Example 89: Colours of light

Question: A photon of wavelength 500 nm is emitted by a traffic light.

1. What is the energy of the photon?
2. What is the frequency of the photon?
3. Use table 16.2 to determine the colour of the light.

Answer

Step 1 : What information is given and what do we need to find?

We are given $\lambda = 500 \times 10^{-9}$ m and we need to find the photon's *energy*, *frequency* and *colour*.

Step 2 : Use the equation $E = \frac{hc}{\lambda}$ to find the photon's energy

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{500 \times 10^{-9}} \\ &= 3,98 \times 10^{-19} \text{ J} \end{aligned}$$

The energy of the photon is $3,98 \times 10^{-19}$ J.

Step 3 : We know the energy of the photon, now we can use $E = hf$ to solve for the frequency

$$\begin{aligned} E &= hf \\ f &= \frac{E}{h} \\ &= \frac{3,98 \times 10^{-19}}{6,63 \times 10^{-34}} \\ &= 6 \times 10^{14} \text{ Hz} \end{aligned}$$

The frequency of the photon is 6×10^{14} Hz.

Step 4 : Use the table to find the colour of light

The wavelength given in the question is 500 nm. We can see in the table that green light has wavelengths between 492 - 577 nm. Therefore 500 nm is in this range so the colour of the light is **green**.

**Worked Example 90: Colours and energies of light**

Question: I have some sources which emit light of the following wavelengths:

1. 400 nm,
2. 580 nm,
3. 650 nm,
4. 300 nm.

What are the colours of light emitted by the sources (see table 16.2)? Which source emits photons with the highest energy and which with the lowest energy?

Answer

Step 1 : What information is given, and what do we need to do?

Four wavelengths of light are given and we need to find their *colours*.

We also need to find which colour photon has the highest energy and which one has the lowest energy.

Step 2 : To find the colours of light, we can compare the wavelengths to those given in table 16.2

1. 400 nm falls into the range for *violet* light (390 - 455 nm).
2. 580 nm falls into the range for *yellow* light (577 - 597 nm).
3. 650 nm falls into the range for *red* light (622 - 780 nm).
4. 300 nm is not shown in the table. However, this wavelength is just a little shorter than the shortest wavelength in the violet range. Therefore 300 nm is *ultraviolet*.

Step 3 : To find the colour of the light whose photons have the highest and lowest energies respectively, we need to calculate the energies of all the photons

We know $E = \frac{hc}{\lambda}$

For 400 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{400 \times 10^{-9}} \\ &= 4,97 \times 10^{-19} \text{ J} \end{aligned}$$

For 580 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{580 \times 10^{-9}} \\ &= 3,43 \times 10^{-19} \text{ J} \end{aligned}$$

For 650 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{650 \times 10^{-9}} \\ &= 3,06 \times 10^{-19} \text{ J} \end{aligned}$$

For 300 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6,63 \times 10^{-34}) \times (3 \times 10^8)}{300 \times 10^{-9}} \\ &= 6,63 \times 10^{-19} \text{ J} \end{aligned}$$

Therefore, the photons with the highest energy are the **ultraviolet** photons.

The photons with the lowest energy are from light which is **red**.

16.4.4 Applications of emission and absorption spectra

The study of spectra from stars and galaxies in astronomy is called *spectroscopy*. Spectroscopy is a tool widely used in astronomy to learn different things about astronomical objects.

Identifying elements in astronomical objects using their spectra

Measuring the spectrum of light from a star can tell astronomers what the star is made of! Since each element emits or absorbs light only at particular wavelengths, astronomers can identify what elements are in the stars from the lines in their spectra. From studying the spectra of many

stars we know that there are many different types of stars which contain different elements and in different amounts.

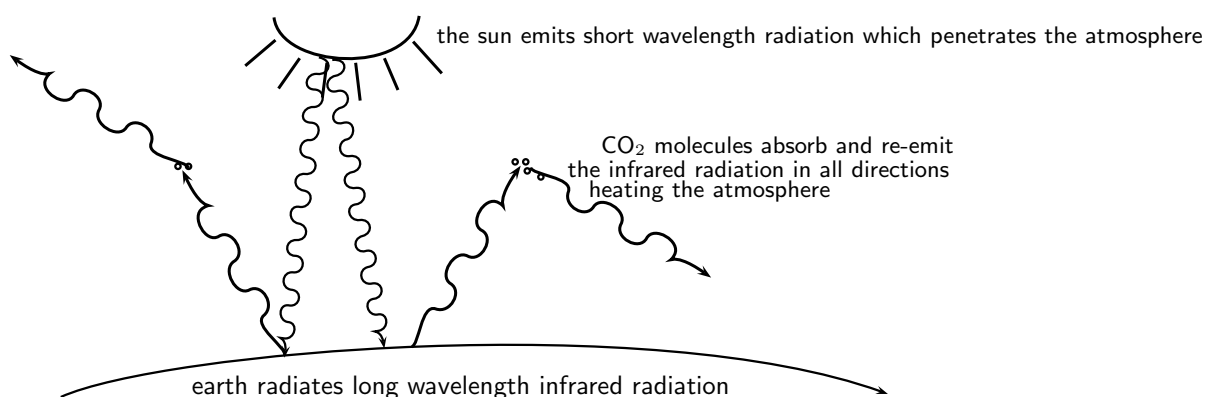
Determining velocities of galaxies using spectroscopy

You have already learned in Chapter 9 about the Doppler effect and how the frequency (and wavelength) of sound waves changes depending on whether the object emitting the sound is moving *towards* or *away* from you. The same thing happens to electromagnetic radiation (light). If the object emitting the light is moving *towards* us, then the wavelength of the light appears shorter (called **blue-shifted**). If the object is moving *away* from us, then the wavelength of its light appears stretched out (called **red-shifted**).

The Doppler effect affects the spectra of objects in space depending on their motion relative to us on the earth. For example, the light from a distant galaxy, which is moving away from us at some velocity, will appear red-shifted. This means that the emission and absorption lines in the galaxy's spectrum will be shifted to a longer wavelength (lower frequency). Knowing where each line in the spectrum would normally be if the galaxy was not moving, and comparing to their red-shifted positions, allows astronomers to precisely measure the velocity of the galaxy relative to the earth!

Global warming and greenhouse gases

The sun emits radiation (light) over a range of wavelengths which are mainly in the visible part of the spectrum. Radiation at these wavelengths passes through the gases of the atmosphere to warm the land and the oceans below. The warm earth then radiates this heat at longer infrared wavelengths. Carbon-dioxide (one of the main greenhouse gases) in the atmosphere has energy levels which correspond to the infrared wavelengths which allow it to absorb the infrared radiation. It then also emits at infrared wavelengths in all directions. This effect stops a large amount of the infrared radiation getting out of the atmosphere, which causes the atmosphere and the earth to heat up. More radiation is coming in than is getting back out.



Therefore increasing the amount of greenhouse gases in the atmosphere increases the amount of trapped infrared radiation and therefore the overall temperature of the earth. The earth is a very sensitive and complicated system upon which life depends and changing the delicate balances of temperature and atmospheric gas content may have disastrous consequences if we are not careful.

Activity :: Investigation : The greenhouse effect

In pairs try to find the following information (e.g. in books, on the internet) and report back to the class in a 5 minute presentation which includes the following:

1. What other gases besides carbon dioxide are responsible for the greenhouse effect?
 2. Where do greenhouse gases come from? (are they human-made or natural?)
 3. Investigate one serious side-effect which could arise if the earth's temperature were to go up significantly. Present some ways in which this effect could be avoided.
-



Exercise: Emission and absorption spectra

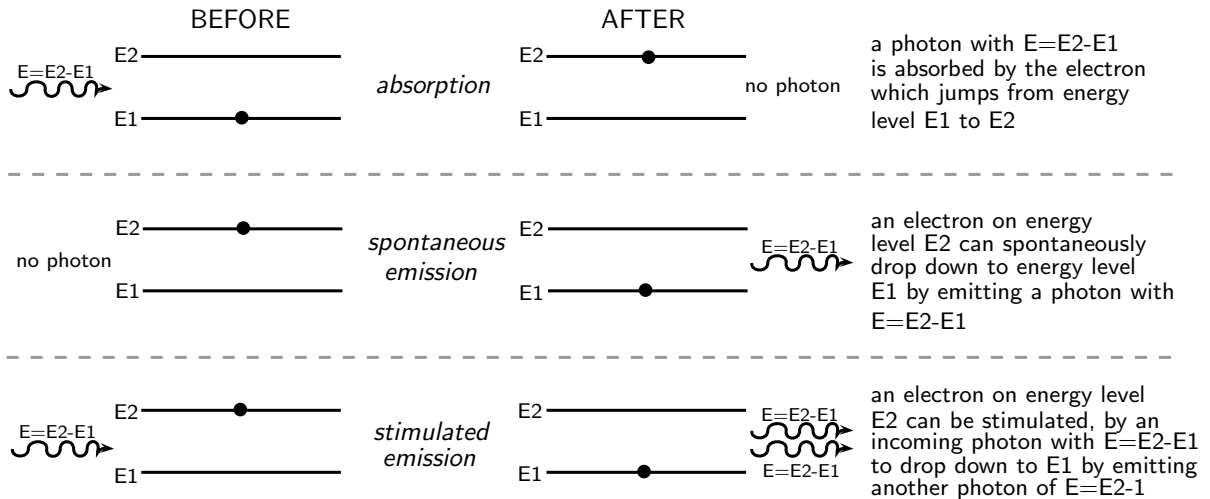
1. Explain how atomic emission spectra arise and how they relate to each element on the periodic table.
 2. How do the lines on the atomic spectrum relate to electron transitions between energy levels?
 3. Explain the difference between atomic absorption and emission spectra.
 4. Describe how the absorption and emission spectra of the gases in the atmosphere give rise to the Greenhouse Effect.
 5. Using table 16.2 calculate the frequency range for yellow light.
 6. What colour is the light emitted by hydrogen when an electron makes the transition from energy level 5 down to energy level 2? (Use figure 16.5 to find the energy of the released photon.)
 7. I have a glass tube filled with hydrogen gas. I shine white light onto the tube. The spectrum I then measure has an absorption line at a wavelength of 474 nm. Between which two energy levels did the transition occur? (Use figure 16.5 in solving the problem.)
-

16.5 Lasers

A laser is a device that produces a special type of light: all the laser photons are identical! They all have the *same* wavelength (and frequency), amplitude and phase. Since they all have the same wavelength, this means they all have the same *colour* and the light is called **monochromatic**. (*Note: mono* means "one" or "single" and *chromatic* means "colour".) This is very different to most other light sources which produce light with a range of wavelengths (e.g. white light from the sun consists of all the visible wavelengths.)

Laser light is highly directional and can be focused very well. This focus allows laser beams to be used over long distances, and to pack a lot of energy into the beam while still requiring reasonably small amounts of energy to be generated. Each centimetre of a typical laser beam contains many billions of photons. These special properties of laser light come from the way in which the laser photons are created and the energy levels of the material that makes up the laser. These properties make laser light extremely useful in many applications from CD players to eye surgery.

The term **LASER** stands for **L**ight **A**mplification by the **S**timulated **E**mission of **R**adiation. This **stimulated emission** is different to the **spontaneous emission** already discussed earlier. Let's review the absorption and emission processes which can occur in atoms.



- **Absorption:** As you can see in the picture above, **absorption** happens when an electron jumps up to a higher energy level by *absorbing* a photon which has an energy equal to the energy difference between the two energy levels.
- **Spontaneous emission:** **Spontaneous emission** is when an electron in a higher energy level drops down to a lower energy level and a photon is emitted with an energy equal to the energy difference between the two levels. There is no interference in this process from outside factors. Usually spontaneous emission happens very quickly after an electron gets into an excited state. In other words, the lifetime of the excited state is very short (the electron only stays in the high energy level for a very short time). However, there are some excited states where an electron can remain in the higher energy level for a longer time than usual before dropping down to a lower level. These excited states are called **metastable states**.
- **Stimulated emission:** As the picture above shows, **stimulated emission** happens when a photon with an energy equal to the energy difference between two levels interacts with an electron in the higher level. This *stimulates* the electron to emit an identical photon and drop down to the lower energy level. This process results in two photons at the end.



Definition: Spontaneous Emission

Spontaneous emission occurs when an atom is in an unstable excited state and randomly decays to a less energetic state, emitting a photon to carry off the excess energy. The unstable state decays in a characteristic time, called the lifetime.



Definition: Meta-stable state

A meta-stable state is an excited atomic state that has an unusually long lifetime, compared to the lifetimes of other excited states of that atom. While most excited states have lifetimes measured in microseconds and nanoseconds (10^{-6} s and 10^{-9} s), meta-stable states can have lifetimes of milliseconds (10^{-3} s) or even seconds.



Definition: Stimulated emission

Stimulated emission occurs when a photon interacts with an excited atom, causing the atom to decay and emit another identical photon.

16.5.1 How a laser works

A laser works by a process called *stimulated emission* - as you can tell from what 'laser' stands for! You can imagine that stimulated emission can lead to more and more identical photons being released in the following way: Imagine we have an electron in an excited metastable state and it drops down to the ground state by emitting a photon. If this photon then travels through the material and meets another electron in the metastable excited state this will cause the electron to drop down to the lower energy level and another photon to be emitted. Now there are two photons of the same energy. If these photons then both move through the material and *each* interacts with another electron in a metastable state, this will result in them *each* causing an additional photon to be released. i.e. from 2 photons we then get 4, and so on! This is how laser light is produced.

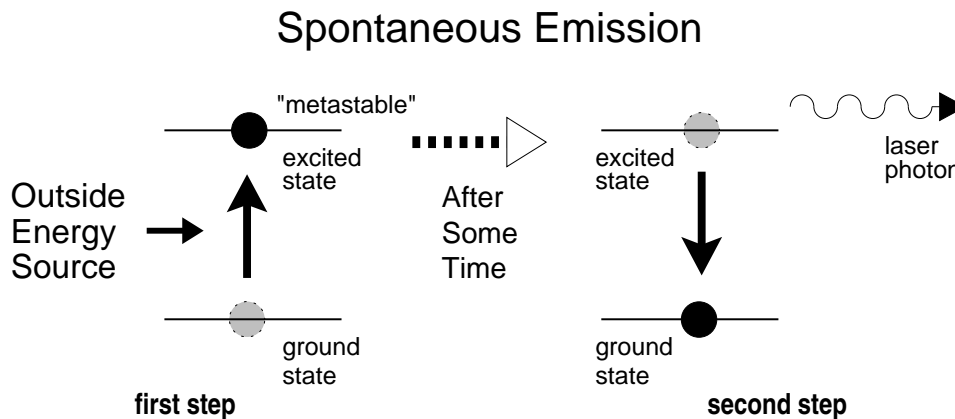


Figure 16.8: Spontaneous emission is a two step process, as shown here. First, energy from an external source is applied to an atom in the laser medium, raising its energy to an excited (metastable) state. After some time, it will decay back down to its ground state and emit the excess energy in the form of a photon. This is the first stage in the formation of a laser beam.

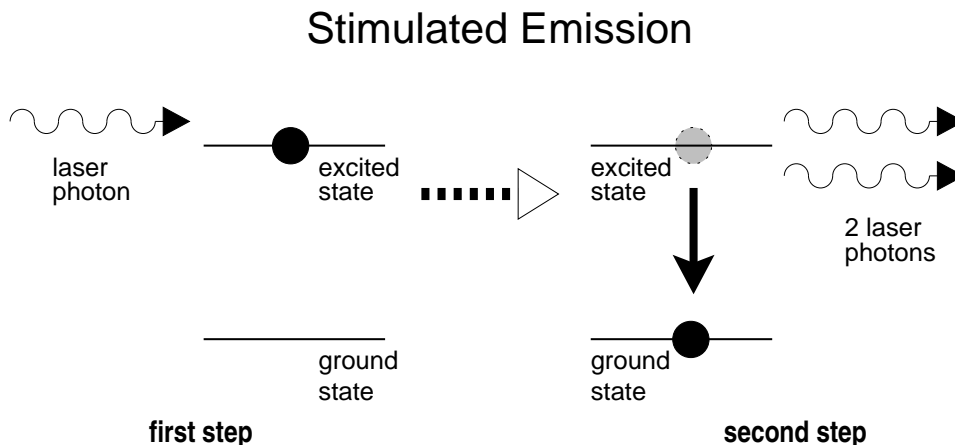


Figure 16.9: Stimulated emission is also a two step process, as shown here. First, a laser photon encounters an atom that has been raised to an excited state, just like in the case of spontaneous emission. The photon then causes the atom to decay to its ground state and emit another photon identical to the incoming photon. This is the second step in the creation of a laser beam. It happens many, many times as the laser photons pass through the optical cavity until the laser beam builds up to full strength.

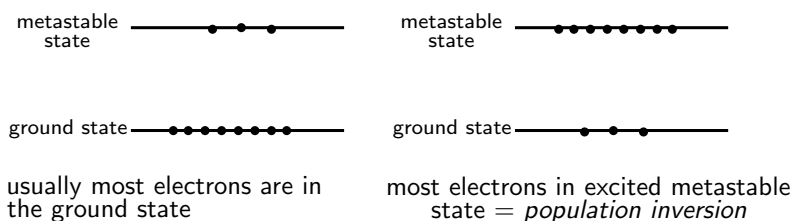
This can only happen if there are many electrons in a metastable state. If most of the electrons are in the ground state, then they will just *absorb* the photons and no extra photons will be emitted. However, if more electrons are in the excited metastable state than in the ground state, then the process of stimulated emission will be able to continue. Usually in atoms, most of the electrons are in the lower energy levels and only a few are in excited states. When most of the

electrons are in the excited metastable state and only a few are in the ground state, this is called **population inversion** (the populations in the excited and ground states are *swapped* around) and this is when stimulated emission can occur. To start off the process, the electrons first have to be excited up into the metastable state. This is done using an external energy source.



Definition: Population inversion

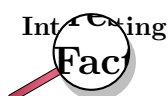
Population inversion is when more atoms are in an excited state than in their ground state. It is a necessary condition to sustain a laser beam, so that there are enough excited atoms that can be stimulated to emit more photons.



Therefore, materials used to make laser light *must* must have metastable states which can allow population inversion to occur when an external energy source is applied. Some substances which are used to make lasers are listed in table 16.3. You can see that gases (such as Helium-Neon mixture), liquids (such as dyes), and solids (such as the precious stone ruby) are all used to make lasers.

Material	Type	Wavelength	Uses
Helium-Neon	gas	632,8 nm	scientific research, holography
Argon ion	gas	488,0 nm	medicine,
Carbon dioxide	gas	10,6 μm	industry (cutting, welding), surgery
Helium-Cadmium	vapor	325 nm	printing, scientific research
Ruby	solid-state	694,3 nm	holography
Neodymium YAG (Yttrium Aluminium Garnet)	solid-state	1,064 μm	industry, surgery, research
Titanium-Sapphire	solid-state	650-1100 nm	research
Laser diode	semiconductor	375-1080 nm	telecommunications, industry, printing, CD players, laser pointers

Table 16.3: A selection of different lasers. The laser material and general type of each laser is given, along with typical wavelengths of the laser light they create. Examples of real-world applications are also given. All these materials allow a population inversion to be set up.



The first working laser, using synthetic ruby as the laser material, was made by Theodore H. Maiman at Hughes Research Laboratories in Malibu, California. Later in the same year the Iranian physicist Ali Javan, together with William Bennet and Donald Herriot, made the first gas laser using helium and neon. Javan received the Albert Einstein Award in 1993.

16.5.2 A simple laser

A laser consists of a number of different parts that work together to create the laser beam. Figure 16.10 shows the different parts of the laser, while Figure 16.11 shows how they create the laser beam.

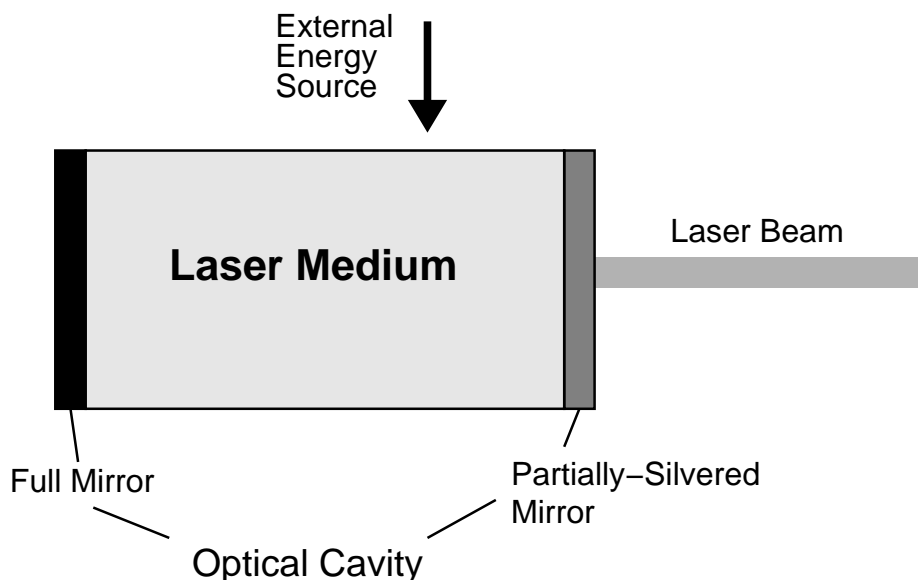


Figure 16.10: Diagram of a laser showing the main components.

The basis of the laser is the laser material which consists of the atoms that are used to create the laser beam. Many different materials can be used as laser material, and their energy levels determine the characteristics of the laser. Some examples of different lasers are shown in Table 16.3. The laser material is contained in the optical cavity.

Before the laser is turned on, all the atoms in the laser material are in their ground state. The first step in creating a laser beam is to add energy to the laser material to raise most of the electrons into an excited metastable state. This is called *pumping* the laser.

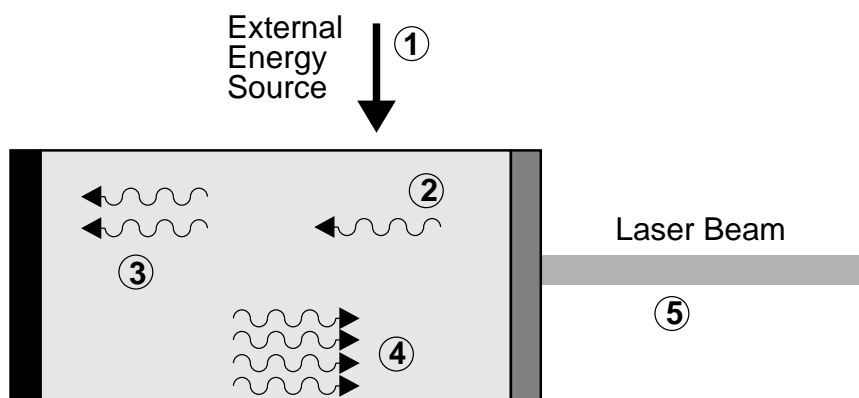


Figure 16.11: Diagram of a laser showing the process of creating a laser beam. (1) A source of external energy is applied to the laser medium, raising the atoms to an excited state. (2) An excited atom decays through spontaneous emission, emitting a photon. (3) The photon encounters another excited atom and causes it to decay through stimulated emission, creating another photon. (4) The photons bounce back and forth through the laser medium between the mirrors, building up more and more photons. (5) A small percentage of the photons pass through the partially-silvered mirror to become the laser beam we see.

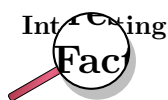
The creation of the laser beam starts through the process of *spontaneous emission*, shown in Figure 16.8. An electron drops down to the ground state and emits a photon with energy equal to the energy difference of the two energy levels. This laser photon is the beginning of the laser beam.

At some time a laser photon will run into another excited electron. Then stimulated emission occurs and the electron drops down to the ground state and emits an additional identical photon as shown in Figure 16.9. Since the laser material typically has a large number of atoms, one laser photon passing through this material will rapidly cause a large number of photons just like it to be emitted.

The optical cavity keeps the laser photons inside the laser cavity so that they can build up the laser beam. At each end is a concave mirror; one is a full mirror and one is a partial mirror. The full mirror is totally reflective. The partial mirror transmits a small amount of the light that hits it (less than 1%). The mirrors are carefully aligned so that photons that reflect off one mirror become “trapped”, and bounce back and forth between the mirrors many times causing more and more stimulated emission. The photons that eventually escape through the partially-silvered mirror become the laser beam that we see.

As the photons bounce between mirrors, they continually pass through the laser material, stimulating those atoms to emit more photons. This creates an ever increasing beam of photons, all with the same characteristics, all traveling in the same direction. In this way, the optical cavity helps to amplify the original laser photons into a concentrated, intense beam of photons.

The laser cavity also helps to narrow the frequency range of laser light emitted. The distance between the two mirrors defines the cavity *mode* which only allows light of a narrow range of frequencies to continue being reflected back and forth. Light of other frequencies damped out. (This is just like in the chapter on the physics of music where a pipe of a certain length corresponds to a particular wavelength of sound.) Therefore only a narrow frequency of light can be emitted.



In 1953, Charles H. Townes and graduate students James P. Gordon and Herbert J. Zeiger produced the first maser, a device operating on similar principles to the laser, but producing microwave rather than optical radiation. Townes's maser was incapable of making a continuous beam. Nikolay Basov and Aleksandr Prokhorov of the former Soviet Union worked independently and developed a method of making a continuous beam using more than two energy levels. Townes, Basov and Prokhorov shared the Nobel Prize in Physics in 1964.

16.5.3 Laser applications and safety

Although the first working laser was only produced in 1958, lasers are now found in many household items. For example, lasers are well-known through their use as cheap laser pointers. However, lasers can be very dangerous to the human eye since a large amount of energy is focused into a very narrow beam. **NEVER POINT A LASER POINTER INTO SOMEBODY'S EYES - IT CAN BLIND THEM FOREVER.**

Other uses include:

- Semiconductor lasers which are small, efficient and cheap to make are used in CD players.
 - He-Ne Lasers are used in most grocery shops to read in the price of items using their barcodes. This makes the cashiers' job much quicker and easier.
 - High energy lasers are used in medicine as a cutting and welding tool. Eye surgery in particular make use of the precision of lasers to reattach the retinas of patients' eyes. The heat from cutting lasers also helps to stop the bleeding of a wound by burning the edges (called cauterising).
-

Activity :: Case Study : Uses of lasers

Do research in a library or on the Internet on one application of laser technology. Explain how the technology works by using a laser.

You will need to present your findings to the class in the form of a poster. You can think of any useful application, but to give you some ideas of where to start, some applications are listed below:

- laser printers
 - laser communication and fibre optics
 - optical storage
 - using lasers as precision measurement tools
 - your own ideas...
-



Exercise: Lasers

1. Explain what is meant by *spontaneous emission of radiation*.
 2. Explain what is meant by *stimulated emission of radiation*.
 3. List the similarities and differences between spontaneous emission of radiation and stimulated emission of radiation.
 4. How is the light emitted by a laser different from the light emitted by a light bulb?
 5. Describe using a simple diagram, how a laser works. Your description should include the following concepts: metastable state and population inversion.
 6. Give examples of some materials that have been used for lasers. What do all these materials have in common?
 7. Describe how the laser cavity affects:
 - increasing amplification
 - concentrating beam intensity
 - narrowing the frequency of the beam
 8. List some applications of lasers.
-

16.6 Summary

1. Light of the correct frequency can eject electrons from a metal. This is called the photoelectric effect.
2. A metal has a work function which is the minimum energy needed to emit an electron from the metal.
3. Emission spectra are formed by glowing gases. The pattern of the spectra is characteristic of the specific gas.
4. Absorption spectra are formed when certain frequencies of light are absorbed by a material.
5. Lasers are devices that produce a special type of light that has many uses.
6. Lasers have many uses, for example, in CD and DVD players, to cut material, in surgery, in printing, in telecommunications and as laser pointers.

16.7 End of chapter exercise

1. What is the photoelectric effect?
2. Calculate the energy of a photon of red light with a wavelength of 400 nm.
3. Will ultraviolet light with a wavelength of 990 nm be able to emit electrons from a sheet of calcium with a work function of 2,9 eV?
4. What does the acronym LASER stand for?
5. Name three types of lasers and their uses.
6. Write a short essay on the benefits lasers have had on modern society.

Part III

Exercises

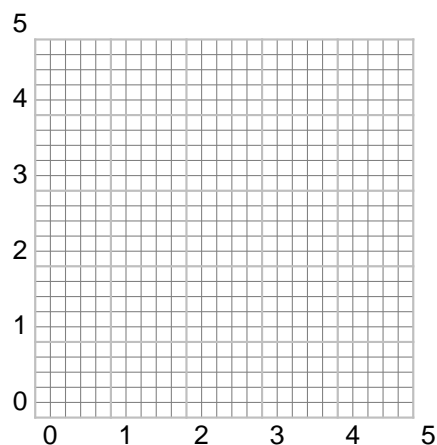
Chapter 17

Exercises

The exercises included in this chapter span over multiple chapters or are no longer required by the syllabus.

- [SC 2002/11 SG] Which one of the following combinations contains one SCALAR and one VECTOR quantity?
 - momentum and force
 - displacement and acceleration
 - potential difference and electric field strength
 - resistance and electric current
- [DOE 2005/11 SG1] Which one of the following pairs consists of two vector quantities?
 - time, acceleration
 - velocity, displacement
 - electric field strength, charge
 - momentum, kinetic energy.
- [IEB 2004/11 HG1] Which one of the following is not equivalent to the SI unit of energy?
 - N.s
 - N.m
 - V.C
 - W.s
- [IEB 2002/11 HG1 - Electrostatics] **Millikan's Oil Drop Experiment**

The diagram below represents the apparatus used to measure the charge carried by oil droplets. The droplets are sprayed above the top plate and eventually a single droplet finds its way through a small hole into the space between the plates.



The mass of the droplet is m and it carries a negative charge of magnitude q . The distance between the plates is d . The oil droplet between the two plates is stationary.

- (a) Use expressions for the electric field intensity and weight to derive a formula for the charge carried by the oil droplet.
 - (b) In one particular experiment, a student reported that, while using a voltage of 400 V between the plates, she had obtained a value of $4,8 \times 10^{-19}$ C for the charge on the oil droplet. Another student said that he had repeated the same experiment with the same oil droplet but that he had used 800 V to keep the oil droplet stationary. If this were true, what would be the charge on the oil droplet, and why would you doubt what he had reported?
5. [IEB 2004/11 HG1] An electric motor lifts a load of mass m through a vertical height h at a steady speed of v . It is connected to a power supply of potential difference V , and it draws a current I . Assume that 90% of the energy input is transferred to useful work done, and that the load comes to rest at the top of this height h .

Which expression correctly relates the power input to the power output for this system?

- (a) $VI = (0,9)mgh$
- (b) $VI = (0,9)mgv$
- (c) $(0,9)VI = mgh$
- (d) $(0,9)VI = mgv$

Part IV

Essays

Essay 1: Energy and electricity. Why the fuss?

Author: Asogan Moodaly

Asogan Moodaly received his Bachelor of Science degree (with honours) in Mechanical Engineering from the University of Natal, Durban in South Africa. For his final year design project he worked on a 3-axis filament winding machine for composite (Glass re-enforced plastic in this case) piping. He worked in Vereeniging, Gauteng at Mine Support Products (a subsidiary of Dorbyl Heavy Engineering) as the design engineer once he graduated. He currently lives in the Vaal Triangle area and is working for Sasol Technology Engineering as a mechanical engineer, ensuring the safety and integrity of equipment installed during projects.

Disclaimer: The details furnished below are very basic and for illustration purposes only.

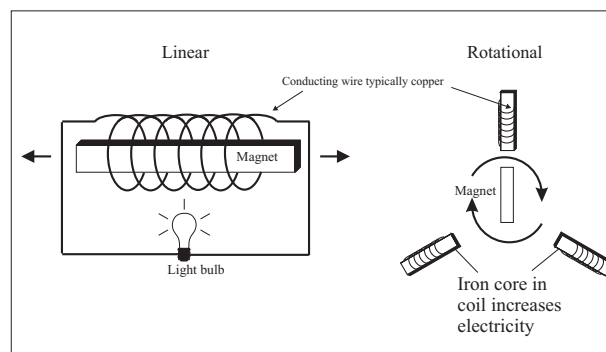
Why do we need energy? Note that I use the word 'energy' and not 'electricity'. On a broad scale it stimulates economic growth, etc, etc but on a personal level it allows us to lead a comfortable lifestyle.

e.g. Flick a switch and

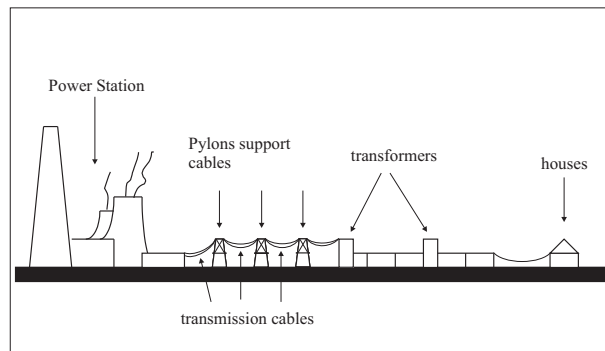
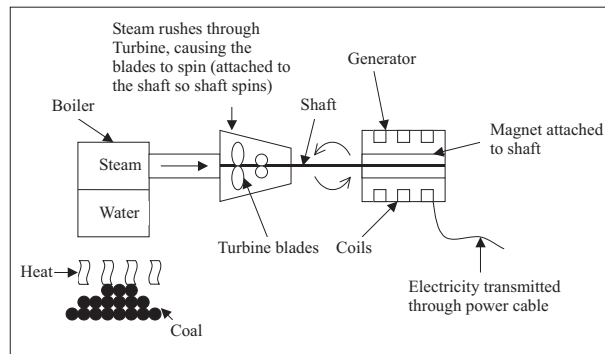
- Heat for cooking
- Entertainment such as television and radio
- Heat for water and interior of house
- Ironing
- Electronic and electrical devices such as alarms, garage doors, etc.

In a modern household this energy is provided in the form of electricity which is powered via fossil fuels or nuclear.

How is electricity made? In a nutshell: By moving a magnet through or near a set of conducting coils.



Most power stations produce steam through heat (nuclear reaction or burning fossil fuels), the steam drives a turbine which moves a magnet relative to a coil (the generator - like the above but on a much larger scale i.e. bigger magnets, bigger coils, etc), which produces electricity that is transmitted via a power network to our homes. Gas fired plants burn gas directly in a gas turbine to produce the same desired relative motion between permanent magnet and coil.



Coal, oil and gas are fossil fuels. Fossil fuels were created by decomposing organic (plant and animal) matter a long, long time ago and are typically found underground. Different temperatures and pressures resulted in the organic matter transforming into coal, oil or gas.

Why the fuss about fossil fuels?

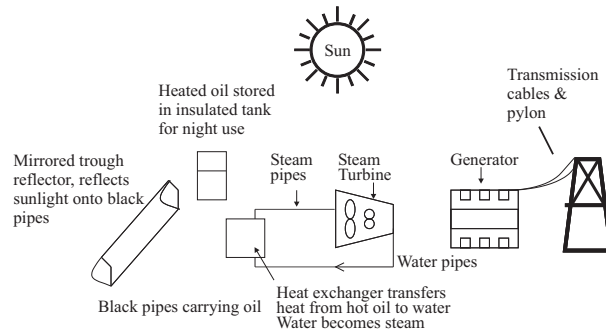
1. Fossil fuel power is bad news in the long run. It pollutes and contributes to the greenhouse effect (global warming resulting in melting polar ice caps, floods, droughts, disease, etc).
2. It's not going to last forever.
3. Nuclear power is 'cleaner' in terms of emissions but there's no proven way of disposing of the nuclear waste. Oh, and it won't last forever either!

Renewable Energy As the name suggests renewable energy lasts 'forever'. Solar (sun), wind, geothermal, wave, hydro and biomass (organic) are all sources of energy that will last until the sun eventually explodes many millions of years from now. Hopefully the human race will have moved from the earth by then! Generally the principal of renewable electricity generation is similar to fossil fuel electricity generation in that electricity is generated by moving a magnet relative to a conducting coil. What is different is the way energy is supplied to cause that motion.

The below are a few different types of available renewable energy technologies.

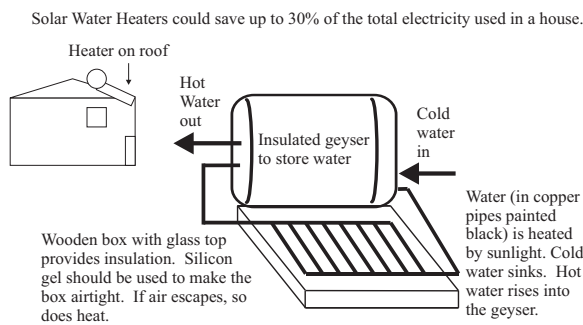
Solar

There are different types of solar electricity technologies, the main ones being solar thermal and photovoltaic.



Solar thermal uses the heat of the sun to produce electricity. Sun is concentrated using mirrors. This heat either creates steam which drives a turbine which in turn drives a generator (as per fossil fuel generation), or drives an air engine (engine that uses expanding air to obtain motion) that drives a generator.

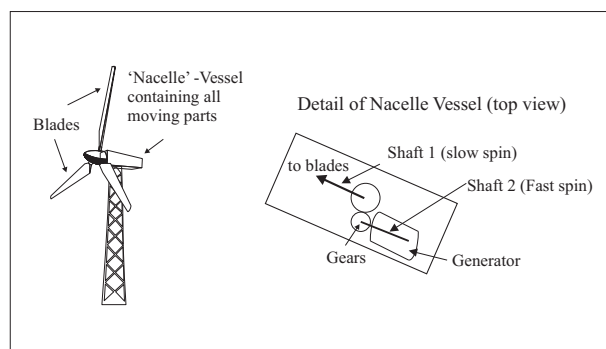
Photovoltaic panels convert sunlight directly into electricity. The benefit of photovoltaic panels is that there are no moving parts, and is therefore relatively maintenance free. The downside is that it's very expensive at this stage (17/06/2004).



Wind

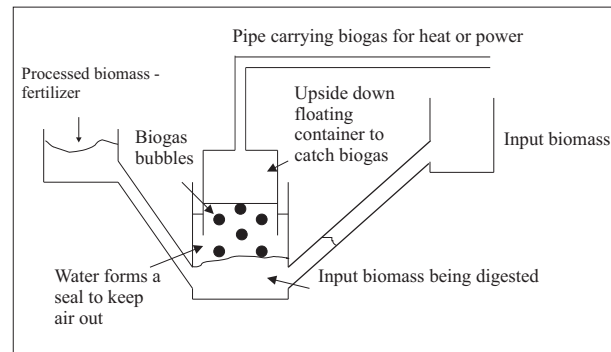
Wind turbines catch wind that spins the blades. The blades are connected to a shaft that spins because of the wind. This spinning shaft spins another shaft that turns a permanent magnet relative to conducting coils.

Note that 'gears' are used to convert the slow spinning of the 1st shaft to a faster spin on the 2nd shaft. The generator shaft needs to spin at the correct speed to produce the right amount and quality of electricity. Some generators are now being modified to run at slower speeds. This saves money as gears are not needed.



Biomass

Biomass is anything organic i.e. plant or animal matter. It can be used in the place of coal as per a normal coal fired plant and is renewable as long as the biomass e.g. wood; is handled in a sustainable manner. By sustainable I mean that suitable farming practices are used so that the land is not 'over farmed' which will result in the soil becoming barren and nothing growing there again.



Biomass can also be processed using 'anaerobic digestion' to produce a gas that can be burned for heat or electricity. This 'biogas' is made up of a number of other gases that are similar to those found in fossil fuel natural gas - Except the amount of the gases are different. E.g. Natural gas has about 94

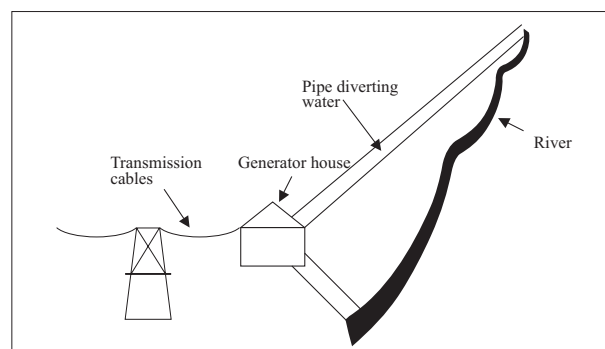
Anaerobic digestion: 'Anaerobic' means 'No air'. Therefore 'anaerobic digestion' means to digest in the absence of air. Bacteria that naturally exist in organic matter will convert organic matter to biogas and fertilizer when all the air is removed.

Thousands of anaerobic digesters have been installed in rural India, Nepal and China in rural area's where cow dung, human waste and chicken litter (faeces) are all processed using anaerobic digestion to produce gas that can be burned in the home for cooking and heating. The leftover is used as fertilizer.

Geothermal Energy

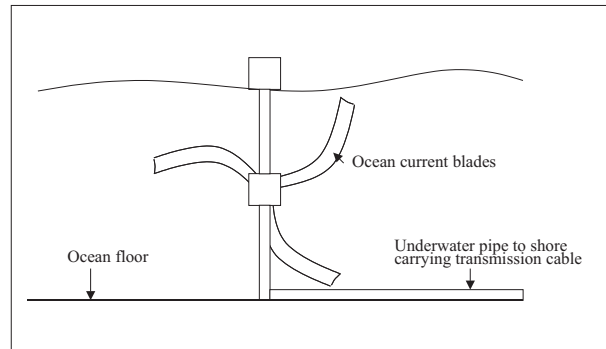
In some places on earth, the earth's crust is thinner than others. As a result the heat from the earth's core escapes. The heat can be captured by converting water to steam, and using the steam to drive a steam generator as discussed above.

Hydroelectric power Water from a river is diverted to turn a water turbine to create electricity similar to the principles of steam generation. The water is returned to the river after driving the turbine.

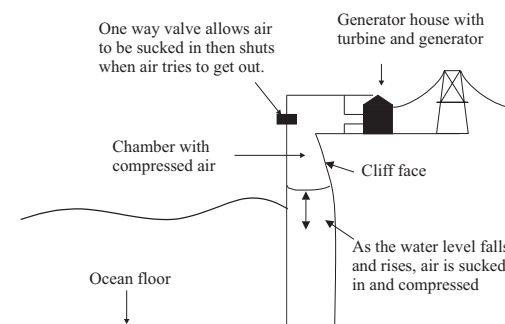


Wave Energy

Some wave energy generators work similarly to wind turbines except that underwater ocean currents turn the blades instead of wind; and of course most of the structure is under water!



Another concept uses the rising and falling of the tides to suck air in using a 'one way valve'. As a result air becomes compressed in a chamber and the compressed air is let out to drive a turbine which in turn drives a generator



These are relatively new technologies.

Liquid Fuels Liquid fuels are used mainly for transportation. Petrol and diesel are the most common liquid fuels and are obtained from oil.

Sasol is the only company in the world that makes liquid fuels from coal; and will be one of the leading companies in the world to make liquid fuels from natural gas! The Sasol petro-chemical plants are based in Sasolburg on the border of the Free State and in Secunda in Mpumalanga.

However, as discussed above coal, gas and oil are fossil fuels and are not renewable. Petrol and diesel are obtained from fossil fuels and therefore pollute and contribute to the green house effect (global warming).

Alternatives

Biodiesel

Oil can be extracted from plants such as the soya bean, sunflower and rapeseed by pressing it through a filter. This oil if mixed correctly with either methanol or dry ethanol and Sodium Hydroxide will separate the plant oil into biodiesel, glycerol and fertilizer.

The biodiesel can be used as produced in a conventional diesel engine with little or no modifications required.

The glycerol can be refined a bit further for pharmaceutical companies to use, or can be used to make soap.

Ethanol

Corn, maize and sugar cane can be used to make ethanol as a fuel substitute for petrol. It's made by the same fermentation process used to make alcohol. Enzymes are often used to speed up the process.

In ethanol from sugar cane production, the leftover 'bagasse' (the fibre part of the sugar cane) can be burned in a biomass power station to produce electricity.

Hydrogen

Through the process of 'electrolysis' electricity (hopefully clean, renewable electricity!) can split water into hydrogen and oxygen. The stored hydrogen can be used in a 'fuel cell' to create electricity in a process that is opposite to electrolysis; to drive electric motors in a car. The hydrogen can also be burned directly in a modified internal combustion engine. In both cases the 'waste' product is water.

Chapter 18

Essay: How a cell phone works

Chapter 19

Essay: How a Physiotherapist uses the Concept of Levers

Chapter 20

Essay: How a Pilot Uses Vectors

How an air traffic controller uses radar

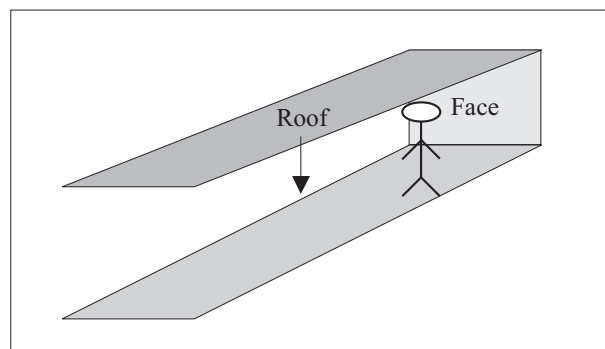
How sonar is used for fishing

Asogan Moodaly received his Bachelor of Science degree (with honours) in Mechanical Engineering from the University of Natal, Durban in South Africa. For his final year design project he worked on a 3-axis filament winding machine for composite (Glass re-enforced plastic in this case) piping. He worked in Vereeniging, Gauteng at Mine Support Products (a subsidiary of Dorbyl Heavy Engineering) as the design engineer once he graduated. He currently lives in the Vaal Triangle area and is working for Sasol Technology Engineering as a mechanical engineer, ensuring the safety and integrity of equipment installed during projects.

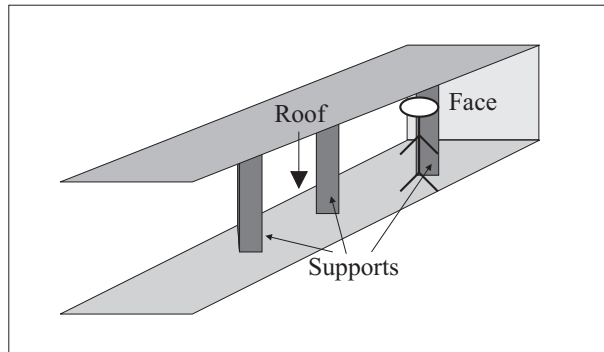
Pressure and Forces

In the mining industry, the roof (hangingwall) tends to drop as the face of the tunnel (stope) is excavated for rock containing gold.

As one can imagine, a roof falling on one's head is not a nice prospect! Therefore the roof needs to be supported.

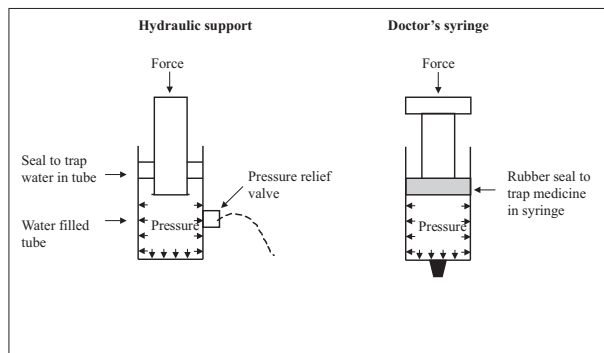


The roof is not one big uniform chunk of rock. Rather it is broken up into smaller chunks. It is assumed that the biggest chunk of rock in the roof has a mass of less than 20 000 kgs therefore each support has to be designed to resist a force related to that mass. The strength of the material (either wood or steel) making up the support is taken into account when working out the minimum required size and thickness of the parts to withstand the force of the roof.

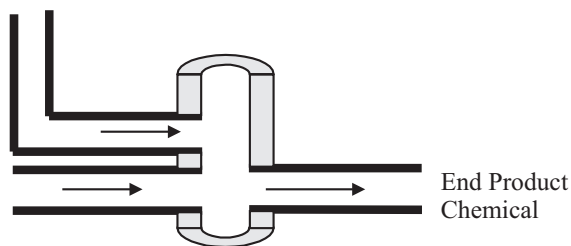


Sometimes the design of the support is such that the support needs to withstand the rock mass without the force breaking the roof..

Therefore hydraulic supports (hydro = water) use the principles of force and pressure such that as a force is exerted on the support, the water pressure increases. A pressure relief valve then squirts out water when the pressure (and thus the force) gets too large. Imagine a very large, modified doctor's syringe.



In the petrochemical industry, there are many vessels and pipes that are under high pressures. A vessel is a containment unit (Imagine a pot without handles, that has the lid welded to the pot that would be a small vessel) where chemicals mix and react to form other chemicals, amongst other uses.



The end product chemicals are sold to companies that use these chemicals to make shampoo, dishwashing liquid, plastic containers, fertilizer, etc. Anyway, some of these chemical reactions require high temperatures and pressures in order to work. These pressures result in forces being applied to the insides of the vessels and pipes. Therefore the minimum thickness of the pipe and vessels walls must be determined using calculations, to withstand these forces. These calculations take into account the strength of the material (typically steel, plastic or composite), the diameter and of course the pressure inside the equipment. Let examine the concepts of force and pressure in further detail.

Appendix A

GNU Free Documentation License

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.

59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary

Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, \LaTeX input format, SGML or XML using a publicly available DTD and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or non-commercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections A and A above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

8. Include an unaltered copy of this License.
9. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
11. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
13. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.
14. Do not re-title any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.
15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organisation as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section A above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known,

or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section A) to Preserve its Title (section A) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sub-license, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sub-license or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.