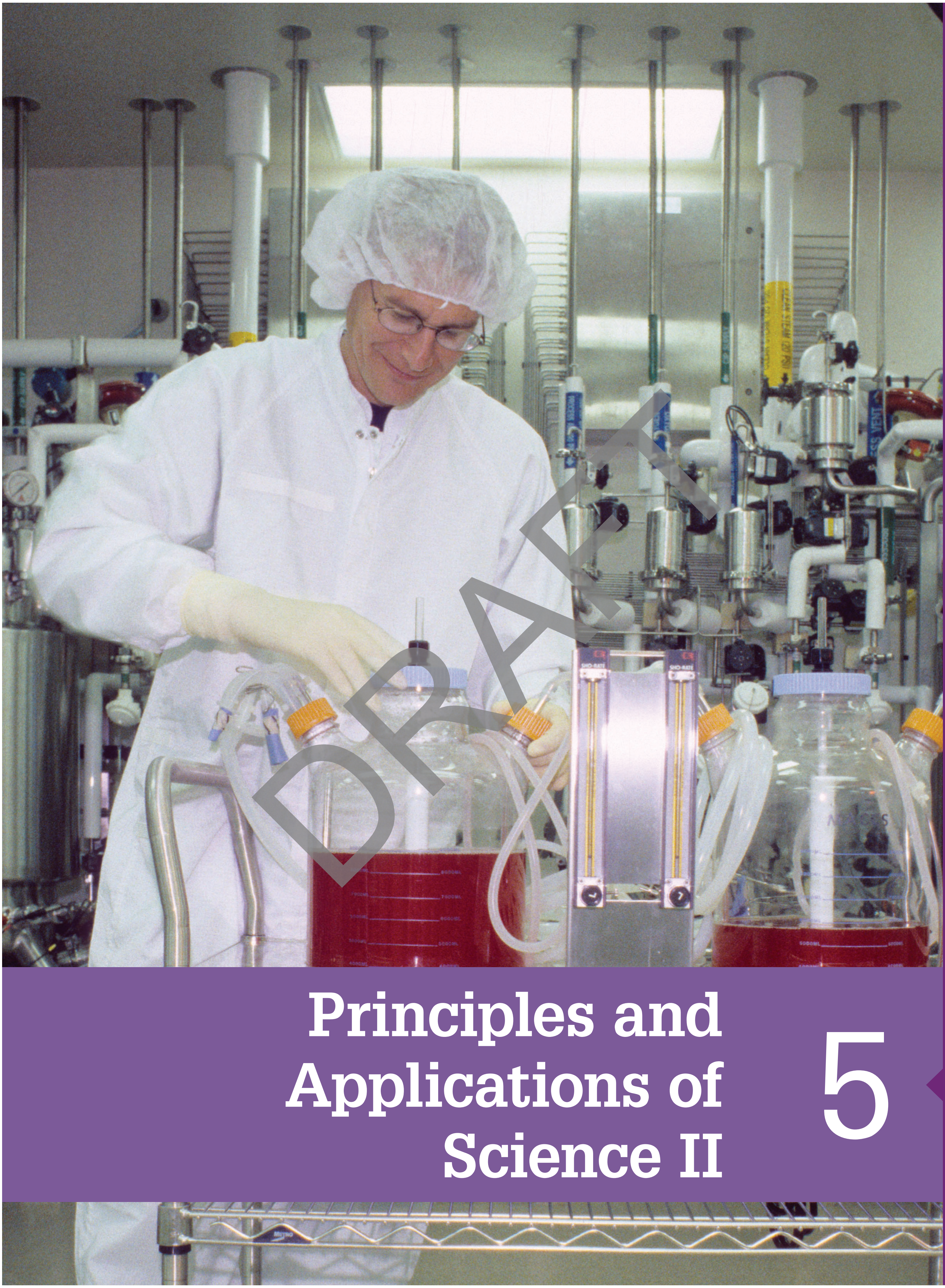


Principles and Applications of Science II

5

Getting to know your unit

Assessment

You will be assessed through a written exam which is set and marked by Pearson.

Science technicians need to be able to use and apply key science concepts to work efficiently and safely in science and science related organisations. Chemists, for example, need to understand how the uses of chemical products relate to their physical and chemical properties. Scientists in the pharmaceutical and medical industries need to understand how the human body works. They need to know how diseases can be diagnosed and treated. Engineers need to understand the properties and behaviour of different materials to ensure that the materials are fit for purpose. When designing machines, they need to understand how materials will behave under different conditions. They also need to understand how energy is transferred in order to make efficient machines and engines.

How you will be assessed

The external assessment for this unit will be split into three sections.

- **Section A** – Chemistry (Properties and uses of substances, Structures, Reactions and properties of commercially important organic compounds, Energy changes in industry)
- **Section B** – Biology (Organs and systems, Ventilation and gas exchange to the lungs, Urinary system structure and function, Cell transport mechanisms)
- **Section C** – Physics (Thermal physics, Materials, Fluids in motion).

The assessment will contain a range of question types, including multiple choice, calculations, short answer and open response. These question types, by their very nature, generally assess discrete knowledge and understanding of the content in this unit.

You need to be able to apply and synthesise knowledge from this unit. The questions will be contextualised in order for you to show that you can do this.

Throughout this unit you will find assessment practices that will help you prepare for the assessment. Completing each of these will give you an insight into the types of questions that will be asked and, importantly, how to answer them.

Unit 5 has four Assessment Outcomes (AOs). These are:

- **AO1**: Demonstrate knowledge of scientific facts, terms, definitions and scientific formulae
  - Command words: describe, draw, explain, identify, name, state
- **AO2**: Demonstrate understanding of scientific concepts, procedures, processes and techniques and their application
  - Command words: calculate, describe, draw, explain, give, show, state
- **AO3**: Analyse, interpret and evaluate scientific information to make judgements and reach conclusions
  - Command words: analyse, comment, describe, explain, give, state
- **AO4**: Make connections, use and integrate different scientific concepts, procedures, processes or techniques
  - Command words: calculate, comment, explain

Command word	Definition – what it is asking you to do
Add/label	This requires you to add labelling to a stimulus material given in the question, for example labelling a diagram or adding units to a table.
Assess	You need to give careful consideration to all the factors or events that apply and identify which are the most important or relevant. You need to make a judgement on the importance of something and come to a conclusion where needed.
Calculate	You will obtain a numerical answer, showing relevant working. If the answer has a unit, you must include it.
Comment on	This requires you to synthesise a number of variables from data/information to form a judgement. More than two factors need to be synthesised.
Compare	This asks you to identify the main factors of two or more items and point out their similarities and differences. You may need to say which are the least important or most important. The word *Contrast* is very similar.
Complete	This requires you to complete a table/diagram.
Criticise	Here you need to inspect a set of data, an experimental plan or a scientific statement and consider the elements. Look at the merits and/or faults of the information presented and back up the judgements that you make.
Deduce	Here you draw/reach conclusion(s) from the information provided.
Derive	Here you combine two or more equations or principles to develop a new equation.
Describe	You need to give a full account of all the information, including all the relevant details of any features, of a topic.
Determine	Your answer must have an element that is quantitative from the stimulus provided, or must show how the answer can be reached quantitatively. To gain maximum marks there must be a quantitative element to the answer.
Devise	You plan or invent a procedure from existing principles/ideas.
Discuss	Here you write about the topic in detail, taking into account different ideas and opinions.
Draw	You produce a diagram either using a ruler or using freehand.
Evaluate	Here you bring all the relevant information you have on a topic together and make a judgment on it (for example on its success or importance). Your judgment should be clearly supported by the information you have gathered.
Explain	You make an idea, situation or problem clear to your reader, by describing it in detail, including any relevant data or facts.
Give a reason why	This is when a statement has been made and you only need to give the reasons why.
Give/state/name	These generally require you to recall one or more pieces of information.
Identify	This usually requires you to select some key information from a given stimulus/resource.
Plot	You need to produce a graph by marking points accurately on a grid from data that is provided and then drawing a line of best fit through these points. You must include a suitable scale and appropriately labelled axes if these are not provided in the question.
Predict	You give an expected result.
Show that	You prove that a numerical figure is as stated in the question. The answer must be to at least one more significant figure than the numerical figure in the question.
Sketch	You produce a freehand drawing. For a graph this would need a line and labelled axes with important features indicated. The axes are not scaled.
State and justify/ identify and justify	You make a selection and justify it.
State what is meant by	Here, you give the meaning of a term but there are different ways in which this meaning can be described.
Write	Here the question asks for an equation.

Getting started

This unit builds on prior knowledge. Look at the topics for each section of the exam. Write down two things you already know about each of those topics. Suggest what you are expecting to learn about when studying those topics.

A Properties and uses of substances

Relating properties to uses and production of substances

Scientists need to know the different chemical properties of substances. They can use this knowledge to make new, useful materials.

Metal oxides and metal hydroxides

Alumina is the common name for aluminium oxide. One property of alumina is that it is **amphoteric**. This means that it will act as an **acid** or a **base** depending on the conditions. If an acid is present it will react as a base. If a base is present it will act as an acid.

Key terms

Amphoteric – substance that can act as both an acid and a base.

Acid – a compound containing hydrogen that dissociates in water to form hydrogen ions.

Base – a compound that reacts with an acid to form a salt and water.

When aluminium oxide reacts with an acid it reacts in the same way as other metal oxides, such as magnesium oxide, that are bases.

Aluminium oxide reacts with hot hydrochloric acid to give aluminium chloride and water. The aluminium chloride is soluble so you get aluminium chloride solution.

$$Al_2O_3 + 6HCl \rightarrow 2AlCl_3 + 3H_2O$$

Aluminium oxide will react with bases such as sodium hydroxide to form aluminates. For example:

$$Al_2O_3 + 6NaOH + 3H_2O \rightarrow \underset{\text{(sodium aluminate)}}{2Na_3Al(OH)_6}$$

Other aluminates can form depending on the conditions. Hot concentrated sodium hydroxide will give sodium tetrahydroxoaluminate, $2NaAl(OH)_4$. You will not be expected to know all aluminates that may form.

In general, aluminium oxide is chemically inert except under certain conditions, e.g. when a hot acid or base is present. This means it has a lot of uses including filler, paint, sunscreens, and glass. The amphoteric nature of alumina means it can be used as a medium for chromatography as a basic, acidic or neutral medium.

It is also an effective desiccant and can be used at a range of basic and acidic pHs. It is stable over pH range 2–13. This also makes it suitable for environmental clean-up and separation applications.

Group 1 and group 2 metal oxides are basic. If they dissolve in water, then they will react with the water to form metal hydroxides. This forms an **alkali** solution. For example:

$$Na_2O + H_2O \rightarrow 2NaOH$$

$$CaO + H_2O \rightarrow Ca(OH)_2$$

Key term

Alkali – a base that dissolves in water to form hydroxide ions.

The oxide ion is a very strongly basic anion due to its very small size and high (2−) charge. The oxide ion in the metal oxide reacts with water to produce hydroxide ions, because a hydroxide ion is the strongest base that can persist in water.

Calcium oxide (quicklime) reacts with water to form calcium hydroxide (lime). This can then be used by farmers to raise the pH of acidic soil.

Magnesium oxide can be used as a desiccant when preserving books in a library. Some papers produce acidic sulfur oxides when oxidised in air. Although magnesium oxide is not a very good desiccant, it is used because due to its basic nature it will neutralise the acidic sulfur oxides produced and so help to preserve the book.

Sodium hydroxide, which is produced when sodium oxide reacts with water, is an important compound in the chemical industry as well as in everyday life. For example, it is used in some processes to make plastics and soaps. It is used in food processing in many ways. Some examples are to peel fruits and to process cocoa and chocolate. It is found in drain cleaner or in oven cleaner. These uses all rely on the basic properties of the compound.

Magnesium hydroxide, $Mg(OH)_2$, is also known as 'milk of magnesia' and is used for treating acid indigestion. It raises the stomach pH due to its basic nature.

Calcium hydroxide, $Ca(OH)_2$, is used in the treatment of acidic effluent. Factories that produce or use sulfuric acid will also produce acidic liquid effluent which may have to be treated before being released to the environment, especially if it has high acid content. Impurities are precipitated and removed. The acidic liquid can be neutralised using calcium hydroxide, lime.

This reaction occurs.

$$H_2SO_4 + Ca(OH)_2 \rightarrow CaSO_4 + 2H_2O$$

The calcium hydroxide is a base and a neutral calcium sulfate is formed. Calcium sulfate is a **salt**. Although other bases are available, calcium hydroxide is cheap and simple to use.

Key term

Salt – a compound formed by an acid–base reaction where the hydrogen in the acid has been replaced by a metal (or other positive) ion.

Electrolysis

Electrolysis breaks down compounds into simpler substances. It is used in some industrial processes (see page 9). Ionic substances are decomposed by an electrical charge being passed through them during electrolysis. The ions must be free to move for this to work so the compound must either be molten (reduced to liquid form by heating) or in solution.

Key term

Electrolysis – the decomposition of a compound using electricity.

Figure 5.1 shows the electrolysis of molten sodium chloride. Sodium is produced at the cathode (positive electrode) and chlorine is produced at the anode (negative electrode).

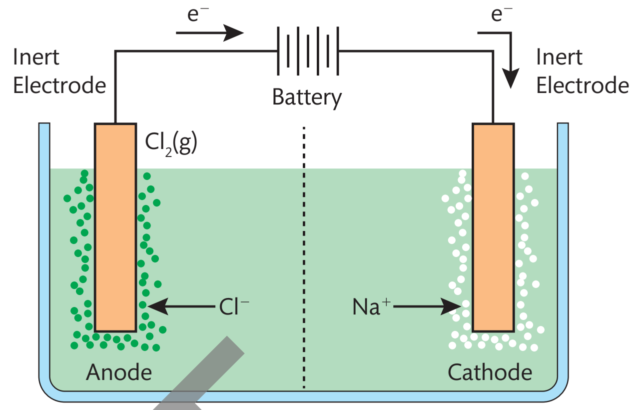


Figure 5.1 Electrolysis of molten sodium chloride

The only ions present are sodium and chloride ions and so only sodium and chlorine can be produced by this reaction.

Figure 5.2 shows the electrolysis of sodium chloride solution.

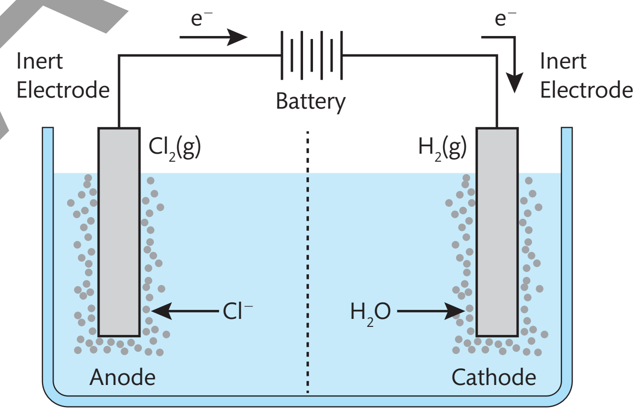


Figure 5.2 Electrolysis of sodium chloride solution

When the sodium chloride is dissolved in water, hydrogen and hydroxide ions are present as well as the sodium and chloride ions. In this case, the chlorine is still produced at the anode but hydrogen gas is produced at the cathode rather than sodium. It is only possible for one type of ion to discharge at each electrode. The ion which is selected to discharge at an electrode depends on a number of factors.

The position of the ion in the electrochemical series affects its ease of discharge at an electrode. When metals react, they lose electrons to become positive ions. If a metal is placed in water, then the metal atoms will tend to lose electrons to the water and become positive. This in turn attracts the negative electrons back to the metal. This is an example of an equilibrium process.

For example, magnesium in water will have the following equilibrium reaction.

$$\mathrm{Mg^{2+}(aq) + 2e^- \rightleftharpoons Mg(s)}$$

A piece of copper in water will behave in the same way.

$$\mathrm{Cu^{2+}(aq) + 2e^- \rightleftharpoons Cu(s)}$$

As copper is less reactive than magnesium, it will form ions less easily so the equilibrium for copper in water is further to the left than that for magnesium in water.

For both the metals, there is a potential difference between the negative charge on the metal and the positive charge of the solution around it. The potential difference is bigger for magnesium than it is for copper. This potential difference can be measured as a standard electrode potential and then the metals can be placed in order of the standard electrode potentials (E°) in the electrochemical series.

Hydrogen is also included in the electrochemical series (see Table 5.1).

Table 5.1 Part of the electrochemical series

Equilibrium half equation	**E° (volts)**
$\mathrm{Li^+(aq) + e^- \rightleftharpoons Li(s)}$	−3.03
$\mathrm{K^+(aq) + e^- \rightleftharpoons K(s)}$	−2.92
$\mathrm{Na^+(aq) + e^- \rightleftharpoons Na(s)}$	−2.71
$\mathrm{Mg^{2+}(aq) + 2e^- \rightleftharpoons Mg(s)}$	−2.37
$\mathrm{Al^{3+}(aq) + 3e^- \rightleftharpoons Al(s)}$	−1.66
$\mathrm{Zn^{2+}(aq) + 2e^- \rightleftharpoons Zn(s)}$	−0.76
$\mathrm{2H^+(aq) + 2e^- \rightleftharpoons H_2(g)}$	0
$\mathrm{Cu^{2+}(aq) + 2e^- \rightleftharpoons Cu(s)}$	+0.34
$\mathrm{\tfrac{1}{2}Cl_2(aq) + e^- \rightarrow Cl^-(aq)}$	+1.36

The lower the metal is in the electrochemical series, the more likely it is to be discharged. So in a solution of copper sulfate, the copper ions will be discharged to form copper atoms rather than the hydrogen ions. In a solution of sodium chloride, hydrogen gas is given off at the cathode rather than sodium metal.

The concentration of the ions in solution will also have an effect on which ion is discharged. The most concentrated ion tends to be discharged, no matter where the ions are in the electrochemical series. So in a concentrated solution containing the chloride ion, the chloride ion will be discharged rather than the hydroxide ion, even though the hydroxide ion is more easily oxidised.

Assessment practice 5.1

Explain the products of the electrolysis of brine.

Transition metals

Transition metals are the *d* block elements found between group 2 and 3 on the periodic table. They are used in a variety of industrial processes and transition metals and their compounds have a wide number of uses. Transition metals have incomplete *d*-sub-shells as a stable ion.

If we look at the transition elements in period 4, across the period, from scandium to zinc, the 3*d*-orbitals are being filled. The pattern is regular. (Except for chromium and copper, which do not follow the principle of completely filling the lowest energy levels first.) The sub-shell energy levels in the third and fourth energy levels overlap. The 4*s*-sub-shell fills before the 3*d*-sub-shell.

When they react, transition element atoms lose electrons to form positive ions. Transition metals lose their 4*s* electrons before their 3*d* electrons. They form ions with more than one stable oxidation state. They can all form compounds with metal ions in the +2 oxidation state.

In solution, transition metal compounds form **complex ions**. A complex ion consists of a central metal ion surrounded by **ligands**. A ligand is a molecule or ion that donates a pair of electrons to the central transition metal ion to form a dative covalent bond.

Figure 5.3 shows the complex ion $\mathrm{[Fe(H_2O)_6]^{2+}}$. Each of the six $\mathrm{H_2O}$ ligands forms one dative covalent bond with the central metal ion.

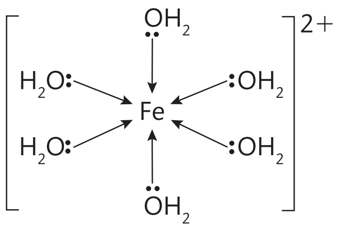


Figure 5.3 Complex metal ion

Other complex ions include $\mathrm{[Cu(H_2O)_6]^{2+}}$, $\mathrm{[Al(H_2O)_6]^{3+}}$, and $\mathrm{[CuCl_4]^{2-}}$.

Transition metals, oxides of transition metals and transition metal complexes are all used as industrial **catalysts**.

Key terms

Complex ion – a transition metal ion bonded to one or more ligands by dative covalent bonds.

Ligand – a molecule or ion that can donate a pair of electrons to the transition metal ion to form a dative bond.

Catalyst – a substance that speeds up a reaction. It can take part in the reaction but is left unchanged at the end of the reaction.

Platinum and rhodium are used in catalytic converters in car exhausts. They convert carbon monoxide and nitrogen monoxide emissions into carbon dioxide and nitrogen gas. These gases are not polluting. Manganese dioxide acts as a catalyst in the decomposition of hydrogen peroxide into water and oxygen.

The contact process uses vanadium(V) oxide as a catalyst for the production of sulfuric acid. There are four stages in this process.

1 Sulfur reacts with oxygen to produce sulfur dioxide.

$$S(s) + O_2(g) \rightarrow SO_2(g)$$

Or sulfur dioxide is produced by reacting sulfide ores with excess air.

$$4FeS_2(g) + 11O_2 \rightarrow 2Fe_2O_2(s) + 8SO_2(g)$$

The sulfur dioxide produced is mixed with excess air for use in stage 2.

2 Sulfur dioxide is converted into sulfur trioxide in a reversible reaction.

$$2SO_2(g) + O_2(g) \rightleftharpoons 2SO_3(g)$$

Vanadium(V) oxide catalyses this stage. The following reactions occur.

The overall reaction is:

$$SO_2 + \tfrac{1}{2}O_2 \xrightarrow{V_2O_5} SO_3$$

Vanadium(V) oxide can act as a catalyst due to the ability of vanadium to change its oxidation state. Sulfur dioxide is oxidised to sulfur trioxide by the vanadium(V) oxide. So the vanadium(V) oxide is reduced to vanadium(IV) oxide.

$$SO_2 + V_2O_5 \rightarrow 2SO_3 + V_2O_4$$

The vanadium(IV) oxide is then oxidised with the oxygen:

$$V_2O_4 + \tfrac{1}{2}O_2 \rightarrow V_2O_5$$

So the vanadium(V) oxide catalyst takes part in the reaction and is changed in the reaction but remains chemically unchanged by the end of the reaction.

3 The sulfur trioxide is converted into sulfuric acid. Adding sulfur trioxide to water is too uncontrollable to be the method used. First sulfur trioxide is dissolved in concentrated sulfuric acid. This produces fuming sulfuric acid which can then be added safely to water to produce more concentrated sulfuric acid.

$$H_2SO_4(l) + SO_3 \rightarrow H_2S_2O_7(l)$$
$$H_2S_2O_7(l) + H_2O(l) \rightarrow 2H_2SO_4(l)$$

Iron is used as the catalyst in the Haber process for the synthesis of ammonia. Nitrogen reacts with hydrogen to produce ammonia.

$$N_2(g) + 3H_2(g) \rightleftharpoons 2NH_3(g)$$

The iron catalyst speeds up the reaction by lowering the **activation energy** so that the N_2 bonds and H_2 bonds can be broken more readily.

Key term

Activation energy – the minimum amount of energy needed by the reactants for collisions to result in a reaction taking place.

Case study

The Haber Process

The iron catalyst lowers the energy demands of the Haber Process. This means that the costs of the process are reduced and also helps the environment. Less fuel is needed to be burned to generate the energy needed. This means fossil fuels are conserved and there are fewer carbon dioxide emissions. The Haber Process is an important industrial reaction because it produces ammonia. This is used as the basis for making fertilisers which improve crop yield. This becomes more important each year as we need to feed the world's population, which is increasing every year.

Check your understanding

1 How does the use of iron in the Haber Process lower environmental and financial costs?

PAUSE POINT List all the different chemical substances discussed so far in this unit.

Hint Compete a table with the headings metals, metal oxides and metal hydroxides.

Extend Give one use for each of the substances based on its chemical properties.

Assessment practice 5.2

Describe how vanadium(V) oxide acts as a catalyst in the contact process. Include any relevant equations.

Purification, extraction and manufacture of useful substances

Alumina (aluminium oxide) is a material that retains its strength at high temperatures. It is chemically and physically stable at these high temperatures. It is a **refractory material**. It can be used in linings for furnaces, kilns and reactors. It can also be used in acidic reactions as it does not react.

Key term

Refractory material – material that is physically and chemically stable at very high temperatures, for example, over 3000 °C.

Alumina and aluminium extraction

Alumina is extracted from **bauxite** ore using the Bayer process. The bauxite is crushed to form grains. It is then mixed with liquor from the later precipitation stage and crushed again to make a **slurry**. Bauxite can have a high level of silica present so it may have to go through a desilication process to remove this. Hot caustic soda (NaOH) solution is used to dissolve the aluminium-bearing minerals, gibbsite, böhmite and diaspore, in the bauxite, to form a sodium aluminate supersaturated solution (liquor). This is called 'digestion'.

Key terms

Bauxite – aluminium ore.

Slurry – semi-liquid mixture containing fine solid particles.

Conditions for this change due to the composition of the bauxite ores. Those with a high gibbsite content need about 140 °C, while those with a high böhmitic content need a higher temperature between 200 and 280 °C.

After digestion, the slurry is cooled to around 106 °C. It is then clarified to separate the solids from the liquor via sedimentation. Chemical additives, **flocculants**, are added to help the **sedimentation** process. The bauxite residue is transferred to the washing tanks to recover the caustic soda, which is reused in the digestion process.

Key terms

Flocculants – substances that causes particles to clump and so settle out of a liquid.

Sedimentation – small solid particles settling at the bottom of a liquid.

Further separation of the pregnant liquor (the sodium aluminate remains in the solution) from the bauxite residue is performed using a series of filters to ensure that the final product is not contaminated with impurities present in the residue.

Precipitation is used to recover the alumina by crystallisation from the liquor, which is supersaturated in sodium aluminate. The liquor is cooled, resulting in the formation of small crystals of aluminium trihydroxide ($Al(OH)_3$), which then grow and form larger crystals.

The spent liquor is heated through a series of heat exchangers and subsequently cooled in a series of flash tanks. The **condensate** formed in the heaters is re-used in the process, e.g. for washing bauxite residue. The remaining caustic soda is washed and recycled back into the digestion process.

The gibbsite crystals formed in precipitation are classified into size ranges. This is normally done using cyclones or gravity classification tanks.

The filter cakes are fed into calciners where they are heated to temperatures of up to 1100 °C to drive off moisture. This produces alumina solids. The following equation describes the **calcination** reaction:

$$2Al(OH)_3 \rightarrow Al_2O_3 + 3H_2O$$

Alumina, a white powder, is the product of this step and the final product of the Bayer Process.

Key terms

Condensate – liquid collected by condensation.

Calcination – heating to high temperature to remove free and chemically bonded water.

Aluminium is extracted from alumina using the Hall-Héroult process. Alumina is dissolved in molten cryolite, Na_3AlF_6 (an aluminium ore), and electrolysis is used to extract the aluminium. Aqueous aluminium oxide is not used because the aluminium is readily oxidised by the hydrogen ions present in the solution. Alumina has a melting point of over 2000°C and so it is not practical to use molten alumina. However, if the alumina is dissolved in molten cryolite, then the melting point is a lot lower (about 1000°C) and so electrolysis is possible. Aluminium fluoride is also added to the mixture and this lowers the melting point even further.

The aluminium sinks to the bottom of the cell and is syphoned off. Carbon dioxide is produced at the anode along with hydrogen fluoride from the cryolite.

Electrolysis of brine

Brine is aqueous sodium chloride. As you saw on page 5, the electrolysis of brine produces sodium hydroxide, hydrogen and chlorine, all of which are industrially useful chemicals. Sodium hydroxide is used in food processing and removing pollutants during paper manufacture. Hydrogen is used in the production of hydrochloric acid and as a fuel. Chlorine is a disinfectant and is also used in making plastics. Figure 5.4 shows electrolysis of brine using a Hoffman voltameter.

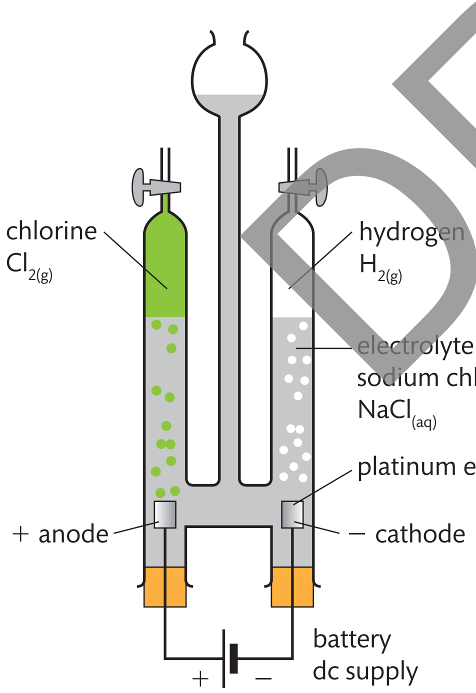


Figure 5.4 Hoffman voltameter

Inert platinum electrodes are used to prevent reaction with the gases produced. Chlorine is collected at the anode and hydrogen is collected at the cathode. The sodium ions react with the hydroxide ions in solution to form sodium hydroxide. The chlorine and the hydrogen gases are kept separate as a mixture of these gases will explode on exposure to heat or light.

The chlorine will also react with the sodium hydroxide solution formed to produce a mixture of sodium chloride and sodium hypochlorate. This mixture can be used as bleach. However, specialised cells are need if the desired products are chlorine and sodium hydroxide. There are two types of cell that can be used:

- a diaphragm
- a membrane cell.

The diaphragm cell has a diaphragm in the centre. This is made of a porous mixture of asbestos and polymers. The brine is added on one side of the diaphragm and the aqueous sodium hydroxide is removed on the other side. The level of liquid on the side where the brine is added is always higher than the side where the aqueous sodium hydroxide is removed, so that the brine does not flow back to mix with the chlorine produced.

The membrane in a membrane cell is made from a polymer that only allows positive ions to pass through it. So the sodium ions can pass through but the chloride ions cannot. So the sodium hydroxide that forms in the right hand compartment does not mix with sodium chloride solution.

The chlorine produced in diaphragm and membrane cells will mix with any oxygen produced at the anode but can be purified by liquefying it under pressure. Under pressure, the oxygen will stay a gas.

Titanium extraction

Titanium can be obtained by extraction from its ore, TiO_2, called rutile. The titanium cannot be extracted using carbon, as titanium carbide (TiC) will form and the presence of this carbide makes the titanium very brittle. There are two stages in the extraction process. It is a batch process so only small amounts are made at one time. It is also expensive due to the two steps, the conditions and the use of chlorine and magnesium. The magnesium has to be extracted from its ore before being used for this process.

1 Conversion of titanium(IV) oxide to titanium(IV) chloride

$$TiO_2(s) + 2Cl_2(g) + 2C(g) \rightarrow TiCl_4(g) + 2CO(g)$$

2 Titanium chloride is then reduced using magnesium to produce titanium.

$$TiCl_4(g) + 2Mg(l) \rightarrow Ti(s) + 2MgCl_2(s)$$

The magnesium is put into a steel reactor and the titanium chloride is pumped in. The reactor is sealed and heated to 1200°C. The reaction is carried out in an argon atmosphere as the titanium produced would react with any oxygen or water present. Any oxygen or nitrogen present in the reactor would make the metal brittle. The reactor is left sealed for two to three days before the titanium can be removed. A large reactor will only produce about 1 tonne of titanium a day, which is a relatively small amount.

It is important to understand that products are used based on the properties they have. These can be physical or chemical properties. Their production is also dependent on their properties. For example, aluminium is very reactive, so is extracted by electrolysis because it cannot be extracted using the cheaper reduction with carbon method. Titanium can be reduced but magnesium is needed to do this as using carbon would cause the metal to be brittle.

PAUSE POINT Draw a mind map of all the production methods discussed in this unit.

Hint Link each to purification, extraction and manufacture methods.

Extend Link each production method to the properties of the substances.

Structures, reactions and properties of commercially important organic compounds

Organic chemicals are carbon based compounds. They are compounds found in living organisms, e.g. amino acids, fatty acids, etc. They are also used in industry and in everyday products. Clothes, plastics, and food are a few examples of where organic chemicals are used.

There are different families of organic compounds. These families are called **homologous series**. The alkanes found in crude oil are a few of the organic compounds found in the alkane homologous series. Alkenes and alcohols are two more homologous series. All members of a homologous series have the same **functional group** and differ only by the number of CH_2 units present. Organic compounds can be classified as **aliphatic**, **alicyclic** or **aromatic**.

Alkanes and alkenes

There are millions of organic compounds. All scientists use the same system to name them. This is the **IUPAC nomenclature**. Scientists follow the rules of this nomenclature to name compounds. Alkanes are aliphatic **saturated compound** hydrocarbons. They only have single bonds and only contain carbon and hydrogen atoms. Methane is an alkane. The 'ane' part of the name shows it is an alkane. The first part of the name, 'meth' shows how many carbons are present in the chain.

Key terms

Homologous series – family of organic chemicals.

Functional group – group of atoms responsible for the characteristic reactions of a substance.

Key terms

Aliphatic – a hydrocarbon with carbon atoms joined together in straight or branched chains.

Alicyclic – a hydrocarbon with carbon atoms joined together in a ring structure.

Aromatic – a hydrocarbon containing at least one benzene ring.

IUPAC nomenclature – system of the International Union of Pure and Applied Chemistry for naming organic compounds.

Saturated compound – contains single bonds only.

The general formula for an alkane is C_nH_{2n+2}. Table 5.2 shows the first ten alkanes, their name, formula and structure.

Table 5.2 Alkanes

Name	Molecular formula
Methane	CH_4
Ethane	C_2H_6
Propane	C_3H_8
Butane	C_4H_{10}
Pentane	C_5H_{12}
Hexane	C_6H_{14}
Heptane	C_7H_{16}
Octane	C_8H_{18}
Nonane	C_9H_{20}
Decane	$C_{10}H_{22}$

Some alkanes are cyclic. These have rings of carbons. For example, cyclohexane has a ring structure with six carbons. Cyclic alkanes have a general formula of C_nH_{2n}.

Alkenes are organic compounds containing a double bond and have the general formula C_nH_{2n}. They have similar names to alkanes with the same number of carbon atoms but the name ends in '–ene'.

Naming organic compounds

Scientists follow these rules when naming an organic compound.

1. The **stem** is the main part of the name and comes from the longest carbon chain or parent chain in the compound.
2. The **suffix** is the end of the name. This identifies the most important functional group, e.g. 'ane' in alkanes.
3. The **prefix** is the front part of the name and identifies other functional groups. The number of the carbon atom they are attached to is given. You count the carbons along the chain so that the functional group has the lowest number possible (e.g. butan-1-ol, butan-2-ol, but-1-ene and but-2-ene).
4. If there is more than one functional group, they are written in alphabetical order.

Table 5.3 shows common functional groups.

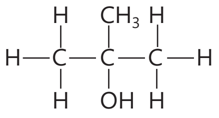
Table 5.3 Common functional groups

Functional group	Formula	Prefix	Suffix
Alcohol	—OH	Hydroxy-	-ol
Aldehyde	—CHO		-al
Alkane	C—C		-ane
Alkene	C=C		-ene
Carboxylic acid	—COOH		-oic acid
Ester	—COO—		-oate
Haloalkane	—F, —Cl, —Br, —I	Fluoro-, chloro-, bromo-, iodo-	
Ketone	C—CO—C		-one

Worked example

Name the following organic compounds.

1



$$\mathrm{H-\underset{H}{\overset{H}{C}}-\underset{OH}{\overset{CH_3}{C}}-\underset{H}{\overset{H}{C}}-H}$$

Stem – the longest chain contains 3 carbon atoms and so the stem is prop-.

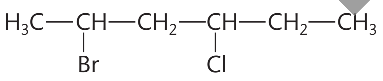
There are two functional groups: i. alcohol —OH, ii. alkyl (methyl-) —CH_3 They are both on the second carbon in the chain.

—OH becomes the suffix. -2-ol

—CH_3 is the prefix 2-methyl-

The name of the compound is 2-methylpropan-2-ol.

2



$$\mathrm{H_3C-\underset{Br}{CH}-CH_2-\underset{Cl}{CH}-CH_2-CH_3}$$

Stem – the longest chain contains 6 carbons so the stem is hex-.

The suffix is –ane, and there are **halo**gens bromine and chlorine present, as this is a **haloalkane**.

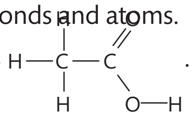
There are two functional groups, bromine and chlorine. These are both prefixes. Remember, if there is more than one prefix they are written in alphabetical order. So bromine is first. This means you count the carbons from the end nearest the bromine as above.

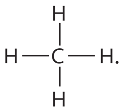
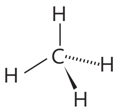
The name of this compound is 2-bromo-4-chlorohexane.

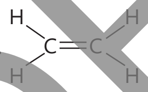
Key term

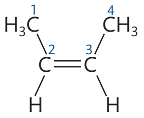
Haloalkane – alkane where at least one hydrogen is replaced by a halogen.

Types of organic formulae

A displayed formula shows the relative positions of all atoms in a molecule and the bonds between them. For example, ethanoic acid has the formula CH_3COOH. Sometimes the functional group is shown as $-COOH$. However, in a displayed formula you must show all the bonds and atoms. The displayed formula for ethanoic acid is  .

Organic compounds are not two-dimensional. It is often important to draw the structure of an **organic compound** to show its three-dimensional structure. Methane can be shown as . This does not show how the hydrogen atoms are arranged in space around the carbon. To see the spatial arrangement, we use wedge and dashed lines.  The solid lines are in line with the plane of the paper. The dashed line shows the bond is angling backwards and the wedge shows the bond is angling forward. It is easier to see the different bond angles when a compound is drawn in this way.

A symmetrical alkene has the same groups attached to each of the carbons in the double bond. Ethene is a symmetrical alkene because both carbons in the double bond are attached to two hydrogens. 

But-2-ene can exist in two distinct geometric forms which have different types of symmetry. Both carbons in the double bond have a methyl group and a hydrogen atom attached.  has a mirror plane symmetry.

But exchanging the positions of the methyl group and the hydrogen attached to carbon atom 3 results in another structure which has a centre of two-fold rotational symmetry.

Isomers

These two forms of but-2-ene are called **isomers**. Isomers are two or more compounds that have the same general formula but a different arrangement of atoms in the compound and so have different properties.

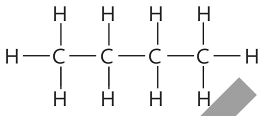
Substances with the same molecular formula but different structural formulae are called structural isomers. There are three types of structural isomer; chain isomers, positional isomers and functional group isomers.

Key terms

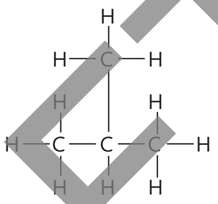
Organic compound – compound based on a carbon structure.

Isomers – two or more compounds that have the same molecular formula but a different arrangement of atoms in the compound and so have different properties.

Butane and 2-methyl propane both have the molecular formula C_4H_{10}. They are chain isomers.



Butane



2-methyl propane

Chain isomers occur in all carbon compounds with four or more carbons. The longer the chain, the more possible chain isomers there are.

Butane and 2-methyl propane are both alkanes so have similar chemical properties but their boiling points are different.

Positional isomers are molecules with the same carbon chain and functional groups but the functional groups are on different carbons within the chain. For example, butanol has the molecular formula $C_4H_{10}O$. It has two positional isomers where the —OH functional group is either on the first or the second carbon in the chain.

Functional group isomers have the same molecular formula but different functional groups, so they belong to different organic families. For example, propanal is an aldehyde with the molecular formula C_3H_6O. Propanone has the same molecular formula but it is a ketone.

Stereoisomerism is caused by a double or triple bond. Optical isomers are also a form of stereoisomerism. Stereoisomers are compounds that have the same structural formula but the arrangement of their atoms in space is different. Atoms can rotate around a single bond. However, this rotation is restricted around a double or triple bond. For example, ethene C_2H_4 has a double $C{=}C$ bond. If a methyl group is added to each carbon, replacing a hydrogen atom, stereoisomers form. These are called cis and trans-isomers. Figure 5.5 shows these stereoisomers.

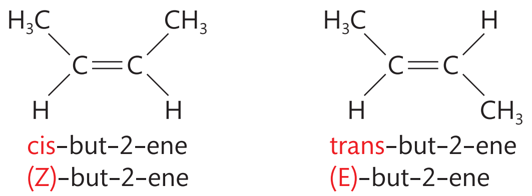


Figure 5.5 (*Z*)-but-2-ene and (*E*)-but-2-ene

E/Z and cis/trans are two different methods of classifying geometric isomers. If something is the cis-isomer, then its name is *cis*-but-2-ene; if it is the E isomer, then its name is (*E*)-but-2-ene. E/Z isomerism is different as it applies a priority to each group attached to the double-bonded carbon and compares the position of those with highest priority. Alkenes showing cis/trans isomerism will also show E/Z isomerism, but the opposite is not true. For example, 1-bromo-1-chloropropene will have E/Z isomerism but not cis/trans isomerism.

Sigma and pi bonds

Carbon has four electrons in its outer shell. These form four covalent bonds by pairing with electrons on other atoms. The electron clouds from each electron in a bond overlap. This is called a sigma bond (σ-bond). C—C bonds and C—H bonds are examples of σ-bonds. The overlap of two *s* orbitals or two *p* orbitals forms σ-bonds.

In alkanes such as methane, the carbon 2*s* and 2*p* atomic orbitals make four hybrid orbitals called sp^3 orbitals. (This notation is a result of hybridisation of the one *s* and three *p* orbitals in the second energy level.) One sp^3 orbital from each carbon overlap to produce a C—C bond. The others overlap with 1*s* orbitals in each of the hydrogen atoms, forming C—H σ-bonds.

The bonding in the double bond in alkenes is slightly different. A σ-bond forms between the two carbon atoms in the same way as in a single bond. A pi-bond (π-bond) is formed by the electrons in adjacent *p*-orbitals overlapping above and below the carbon atoms. The π-bond can only form if a σ-bond has already formed. The π-bond restricts movement around the double bond. This means the region around the double bond in a molecule is flat. Double bonds are 'planar'. The π-bond is the reactive part of the molecule because there is a high electron density around it. Figure 5.6 shows the bonding in ethene.

Bond length and bond shape

Bond length and bond shape in molecules can be explained by the hybridisation of orbitals (see Table 5.4). A bond is formed by the overlap of atomic orbitals. Hybridisation is when atomic orbitals fuse to form newly hybridised orbitals.

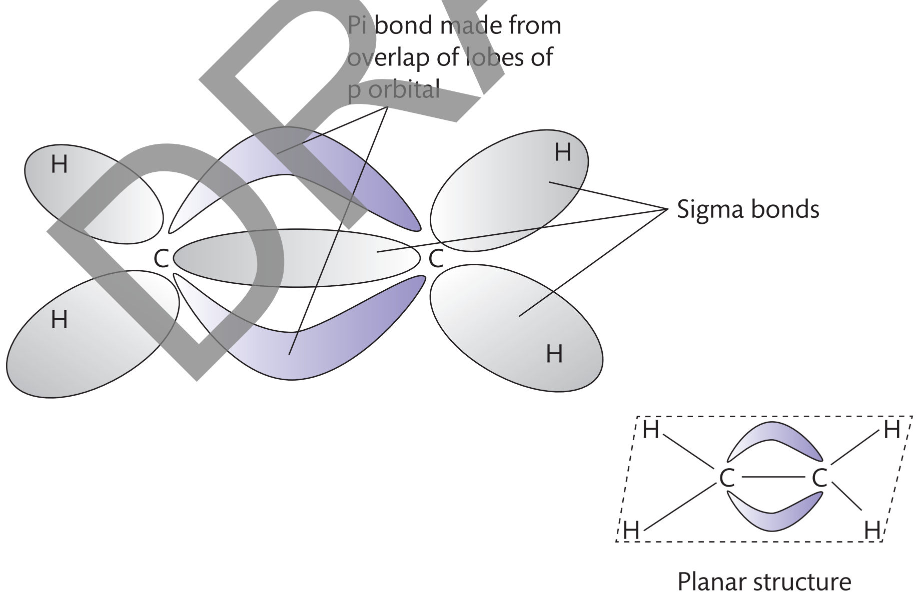


Figure 5.6 Bonding in ethene

Table 5.4 Hybridised orbitals

Hybrid type	Bond angle	Arrangement	Example	C—C bond strength (kJ mol^{-1})	C—C bond length (nm)
sp^3	109.5°	tetrahedral	ethane	347	0.147
sp^2	120°	Trigonal planar	ethene	612	0.135

For example, methane contains four C—H bonds. Carbon has the electronic configuration $1s^2\ 2s^2\ 2p^2$, so only the two *p* electrons are unshared. This would mean only two bonds are possible. However, we know four bonds are formed. The one *s* and three *p* orbitals combine to form hybrid orbitals which have 25% *s* character and 75% *p* character. These are sp^3 hybridised orbitals. This makes all four bonds in methane equivalent to each other, giving equal bond angles of 109° and methane a tetrahedral shape.

In ethene, a π-bond is required for the double bond. This uses sp^2 hybridised orbitals. The 2*s* orbitals mix with two of the three available 2*p* orbitals forming three sp^2 orbitals with a remaining *p* orbital.

The type of bond affects the length and strength of the bond. This is shown in Table 5.5.

Table 5.5 Bond strength and length

Bond	Number of electrons	Bond strength	Bond length
Single	2	Weakest	Longest
Double	4		
Triple	6	Strongest	Shortest

As the number of electrons shared between two atoms increases, the bond strength increases. The distance between the nuclei decreases and so you get a stronger, shorter bond. So the double C═C bond in ethene is stronger and shorter than the single C—C bond in ethane.

The bond strengths and lengths in the benzene ring of an aromatic hydrocarbon are different than those in a straight chain alkene.

The formula for benzene is C_6H_6. All the bond lengths between each carbon are identical. This is because it has a delocalised electron structure. Each carbon atom in benzene donates on electron from its *p*-orbital, forming a ring of delocalised electrons above and below the plane of the molecule. The electrons move freely within the ring and do not belong to any one carbon.

The bonds between the carbons in the benzene ring have a length of 0.140 nm which is between that of a C—C bond and a C═C bond. The bond strength is about 518 kJ mol^{-1}, which is between the bond strengths of the C—C bond and the C═C bond.

In order to show the delocalised electrons in benzene when representing it in an equation, you draw a hexane ring with a circle inside to represent the delocalised electrons, as shown in Figure 5.7.

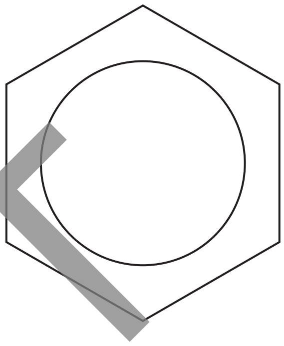


Figure 5.7 The skeletal structure of benzene

Boiling points

As the chain length of the alkanes increases, so does the boiling point. The relative molecular mass also increases as the chain length increases. The larger the molecule, the more surface area is available for contact with adjacent molecules. This means the number of induced dipole-dipole forces increases (van der Waals forces; see Unit 1). So the longer the chain, the stronger the intermolecular attraction and so more energy is needed to overcome this for the molecule to change state.

Structural isomers of alkanes will have different boiling points. This is because they are branched chains and so have a smaller surface area than their straight chain equivalents. Therefore there is less surface area for contact, and so fewer induced dipole-dipole attractions. So branched-chain isomers will have a lower boiling point than the straight-chained isomer.

Pentane has a boiling point of 31.6°C. 2,2-dimethylpropane only has a boiling point of 9.5°C. They are structural isomers of each other.

PAUSE POINT Produce a table showing the name formula and structural formula of the organic compounds discussed in this unit.

Hint You have studied alkanes, alkenes and haloalkanes as well as aromatic compounds.

Extend Choose two different organic molecular formulas and draw out all the isomers for each.

Reactions of organic compounds

Alkanes are not very reactive. This is because a large amount of energy is needed to break the covalent bonds in the molecules. (The bonds have high bond enthalpy– see next page). The electronegativities of carbon and hydrogen are very similar and so the C—H bond is not very polar.

Alkanes are often used as fuels as complete combustion releases large amounts of energy. When there is plenty of air, or excess oxygen, all alkanes will completely combust to form carbon dioxide and water.

The word equation for the combustion of alkanes is:

$$\text{alkane} + \text{oxygen} \rightarrow \text{carbon dioxide} + \text{water}$$

Methane is the fuel used in a Bunsen burner. The equation for the combustion of methane is:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

If oxygen supply is limited, for example, when the hole in the Bunsen burner is closed, incomplete combustion will occur.

$$2CH_4 + 3O_2 \rightarrow 2CO + 4H_2O$$

Many organic reactions take several steps to gain the final product. These steps are called the **reaction mechanism**.

Key term

Reaction mechanism – the step by step sequence of reactions that lead to an overall chemical change occurring. It shows the movement of electrons in the process.

Free radical substitution

Alkanes react with halogens to form haloalkanes. For example, chlorine will react with methane to form chloromethane. The reaction needs ultraviolet light or temperatures of about 300 °C. **Halogenation** happens by **homolytic fission** of the halogen molecule forming **free radicals**. A hydrogen atom in the alkane is substituted with a halogen atom. If the reaction conditions remain then all the hydrogens can be substituted with a halogen.

Key terms

Halogenation – addition of halogen such as bromine or chlorine to an alkane.

Homolytic fission – formation of a free radical by splitting a bond evenly so each free radical has one of the two available electrons.

Free radical – atom with a single unpaired electron.

The free radical substitution mechanism has three stages:

1 Initiation – formation of radicals.
2 Propagation – steps that build up the desired product in a side reaction.
3 Termination – two free radicals collide.

These reactions are unpredictable and difficult to control. A mixture of products is formed and these will need to be separated by fractional distillation or chromatography. The reaction is shown in Table 5.6.

Table 5.6 Free radical substitution

Step	Equation	Description
Initiation	$Cl_2 \rightarrow 2Cl\bullet$	UV light or 300 °C
Propagation	$CH_4 + Cl\bullet \rightarrow$ $\bullet CH_3 + HCl$ $\bullet CH_3 + Cl_2 \rightarrow$ $CH_3Cl + Cl\bullet$	Step 1 generates an alkyl radical and hydrogen chloride. Step 2 generates the desired product and regenerates the chlorine radical.
Termination	$2Cl\bullet \rightarrow Cl_2$ $2\bullet CH_3 \rightarrow C_2H_6$ $\bullet CH_3 + Cl\bullet \rightarrow$ CH_3Cl	A mixture of products is produced by the random collisions between free radicals. The desired CH_3Cl needs to be separated from the rest of the products. Termination is the result of any two radicals colliding.

As there are chlorine radicals present in the mixture, the chloromethane produced could react with these to form other chloroalkanes with chlorine attached to more than one carbon.

Assessment practice 5.3

Write out the three steps for the free radical substitution of ethane, C_2H_6, by bromine, Br.

Electrophilic addition in alkenes

Alkenes are more reactive than alkanes. This is due to the C=C double bond.

The **mean bond enthalpy** for C—C is +347 kJ mol^{-1}. So you might expect the mean bond enthalpy for C=C to be double that, +694 kJ mol^{-1}. In fact, it is only +612 kJ mol^{-1}. This means that the π-bond only uses 265 kJ mol^{-1} of energy to break the bond. The π-bond is weaker than the σ-bond and so breaks more easily and reacts first.

The double bond is an area of high electron density so **electrophiles** such as Br_2 and HBr are attracted to it.

In alkenes, the π-bond breaks and a small molecule is added across the two carbon atoms.

Halogenation is an **addition reaction** in which a halogen is added across the C═C bond. Adding bromine water to an alkene will form a haloalkane and the bromine water will decolourise, because the bromine is removed from the reaction mixture. This is the test for identifying an alkene.

$$CH_2CH_2 + Br_2 \rightarrow CH_2BrCH_2Br$$

Key terms

Electrophilic addition – a reaction using an electrophile where two or more molecules bond to become one product.

Mean bond enthalpy – the average amount of energy for a mole of a given bond to undergo homolytic fission in the gaseous state.

Electrophiles – species that are electron-pair acceptors and are attracted to areas of high electron density.

Addition reaction – a reaction where two or more molecules join together to give a single product.

This is a multi-step process. The reaction mechanism is shown in Table 5.7. The arrows show the movement of the electrons.

Haloalkanes can also be formed by an addition reaction between a hydrogen halide such as HBr and an alkene. Hydrogen halides like HCl, HBr and HI are gases at room temperature so they are bubbled trough the alkene to cause a reaction.

$$CH_2CH_2 + HBr \rightarrow CH_3CH_2Br$$

There is a large electronegativity difference between the hydrogen and the halide in a hydrogen halide, so the molecule is polar. The hydrogen end of the molecule, which is δ^+, is attracted to the high electron density of the double bond. Heterolytic fission occurs in the hydrogen halide and the electrons in the π-bond. A bromine ion and a carbocation then form, and react to form a stable product.

Assessment practice 5.4

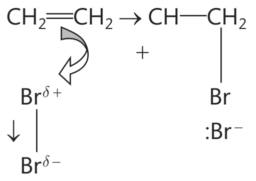
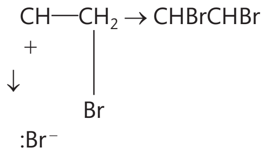
Copy and complete the following table to show the reaction mechanism for the addition of HBr to ethane.

Stage	Mechanism	Explanation

When a hydrogen halide is added to an unsymmetrical alkene there are two possible products. For example, when hydrogen bromide is reacted with propene both 1-bromopropane and 2-bromopropane are formed. 2-bromopropane is the major product formed.

When H-X (X is a halide) is added to an unsymmetrical alkene then the hydrogen becomes attached to the carbon with the most hydrogen atoms to start with. This occurs because carbocations that have alkyl groups attached are more stable than those with only hydrogen atoms attached.

Table 5.7 Halogenation of alkenes

Stage	Mechanism	Explanation
Starting reaction	$CH_2{=}CH_2$ Induced dipole $Br^{\delta+} - Br^{\delta-}$	Bromine has non polar molecules. The electron rich C═C bond induces a dipole in the bromine molecule. The δ^+ end of the bromine molecule is attracted to the high electron density of the C═C bond.
Reaction intermediates	 $CH_2{=}CH_2 \rightarrow \overset{+}{C}H{-}CH_2Br$, $Br^{\delta+}{-}Br^{\delta-}$, $:Br^-$	The electrons from the π-bond make a bond with a bromine atom. This causes heterolytic fission of the bromine molecule. The other carbon atom has a positive charge. This positive carbon atom species is very reactive and is called a carbocation.
Forming dibromoethane	 $\overset{+}{C}H{-}CH_2Br + :Br^- \rightarrow CHBrCHBr$	Two electrons from the bromide ion are shared with the carbocation making a second bond and a stable product.

Hydration is an addition reaction between an alkene and steam. This is when water is added. The alkene must be in gaseous form for the reaction to occur. The temperature must be high, about 300°C, as must be the pressure, about 65 atm. Phosphoric acid acts as the catalyst.

$$CH_2CH_2 + H_2O \rightarrow CH_3CH_2OH$$

Alkenes react with concentrated sulfuric acid under cold conditions to form alkyl hydrogensulfates. For example, ethene reacts to form ethyl hydrogen sulfate.

$$CH_2{=}CH_2 + H_2SO_4 \rightarrow CH_3CH_2HSO_4$$

Sulfuric acid acts as an electrophile in this reaction. The hydrogen atoms are attached to the very electronegative oxygen atoms giving the hydrogen atoms a slight positive charge, δ^+. (This is similar to what happens with the hydrogen atoms in the hydrogen halides.) The positive hydrogen is attracted to the high electron density double bond in the alkene. The mechanism is shown in Figure 5.8.

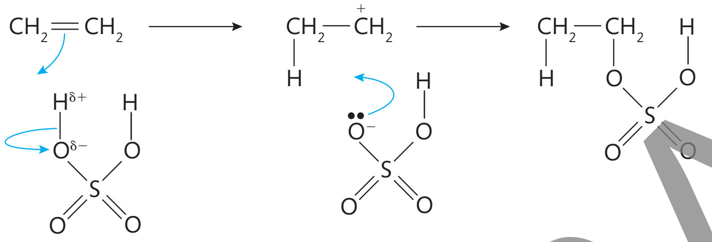


Figure 5.8 Electrophilic addition of sulfuric acid to ethene

Polymer formation

Polymers are important commercial substances and scientists need to understand the reactions and processes involved in producing them.

One method is the free radical polymerisation of alkenes. This is an addition reaction.

Polythene can be produced by an addition reaction. Thousands of ethene molecules join together to make polythene.

$$nNCH_2{=}CH_2 \rightarrow [CH_2{-}CH_2]_n$$

This reaction needs the following conditions: 200°C, 2000 atm, a small amount of oxygen.

- **Initiation:** free radicals are produced by the reaction between ethene and oxygen. A variety of free radicals are produced so it is easier to represent them as Ra.
- **Propagation:** a free radical hits an ethene molecule to form a longer free radical.

$$Ra\bullet + CH_2{=}CH_2 \rightarrow Ra\ CH_2CH_2\bullet$$
$$Ra\ CH_2CH_2\bullet + CH_2{=}CH_2 \rightarrow Ra\ CH_2CH_2CH_2CH_2\bullet$$

and so on

- **Termination:** if two free radicals collide then they will produce a final molecule and no more free radicals are formed.

$$Ra(CH_2)_n\bullet + \bullet(CH_2)_nRa \rightarrow Ra(CH_2)_nRa$$

The randomness of this process means that polythene will be made up of many different length chains.

Many hydrocarbons found in crude oil have very long chains. Many of these hydrocarbons have few uses or are in low demand. Hydrocarbons such as ethene which have short chains are much more in demand. A process called **cracking** is used to decompose long chain hydrocarbons into smaller chain hydrocarbons. There are two types of cracking used in the **petrochemical industry**, thermal and catalytic. The hydrocarbon molecules are broken in a random way during the reaction to produce a mixture of smaller hydrocarbons. Alkanes and alkenes are always produced. For example:

$$C_{10}H_{22} \rightarrow C_5H_{12} + C_3H_6 + C_2H_4$$

Pentane, propene and ethene have been produced here. These are not the only possibilities. Another reaction may produce octane and ethene.

$$C_{10}H_{22} \rightarrow C_8H_{18} + C_2H_4$$

Catalytic cracking uses zeolites as the catalyst. Zeolites are complex aluminasilicates and are large lattices of aluminium, silicon and oxygen atoms. The other conditions needed are a temperature of about 500°C and low pressures. The hydrocarbons are mixed with the catalyst which has been ground into a fine powder. It is an example of homogeneous catalysis, as both the reactants and the catalyst are solids.

This method gives a high number of hydrocarbons with between five and ten carbon atoms in the chain. These are useful for petrol. Catalytic cracking also produces high numbers of branched alkanes and aromatic hydrocarbons such as benzene. The hydrocarbons produced are separated by cooling and fractional distillation.

Thermal cracking uses temperatures between 450°C and 750°C. The pressure can be up to 70 atms. Thermal cracking gives high numbers of alkenes. Higher temperatures mean that the chain breaks near to its end. This gives a high proportion of short chain alkenes.

Key terms

Cracking – long chain hydrocarbons are broken down into shorter chain hydrocarbons.

Petrochemical industry – industry that produces materials by refining petroleum.

PAUSE POINT Describe all the different reaction types in this unit.

Hint Remember to consider the different conditions for each reaction type.

Extend Write out the reaction mechanism for one of the reaction types.

Energy changes in industry

Chemical reactions always involve an energy change. In industry, most products are made on a very large scale. Industrial chemists want to make these products as safely and cheaply as possible. This means they need to also carry out the reactions as quickly as possible with as little energy input as possible.

Enthalpy

Enthalpy H is the thermal energy stored in a chemical system. It is defined by:

$$H = U + pV$$

where U = internal energy of the system

p = pressure

V = volume

You cannot measure the enthalpy of a system but you can measure the energy absorbed or released to the **surroundings** during a reaction. Thermal energy changes are measured by measuring the temperature using the **Kelvin scale**.

Key terms

Surroundings – everything that is not part of the system e.g. water bath or beaker or aqueous solution that reactants are dissolved in.

Kelvin scale – a temperature scale with absolute zero as zero. The size of one unit (1 K) is the same as one degree Celsius. Water freezes at 273 K.

Enthalpy change, ΔH, is the heat exchange with the surroundings during a reaction. It is the difference between the enthalpy of the products and the enthalpy of the reactants. This is at constant pressure.

$$\Delta H = H_{\text{products}} - H_{\text{reactants}}$$

It can also be defined as $\Delta H = \Delta U + p\,\Delta V$.

Enthalpy changes are measured in kJ mol^{-1} (kilojoules per mole). In order to compare enthalpy changes they must be measured in standard conditions. These are a pressure of 1×10^5 Pa and a temperature of 298 K. All substances must be in their standard states. These are their most stable states. When enthalpy changes are measured under standard conditions the enthalpy change is known as ΔH°.

- An **exothemic reaction** releases heat and has a negative enthalpy change ($-\Delta H$). The enthalpy of the products is smaller than that of the reactants. Heat is lost from the system to the surroundings.
- An **endothermic reaction** absorbs heat and has a positive enthalpy change ($+\Delta H$). The enthalpy of the products is greater than that of the reactants. Heat is absorbed by the system from the surroundings. Figure 5.9 shows exothermic and endothermic enthalpy changes.

Key terms

Exothermic reaction – a reaction that releases energy.

Endothermic reaction – a reaction that absorbs energy.

The two diagrams in Figure 5.9 are called enthalpy profile diagrams. In an exothermic reaction, excess energy is released into the surroundings so the temperature of the surroundings will rise. In an endothermic reaction, energy is required. This energy comes from the surroundings so the temperature of the surroundings will fall.

In order for a reaction to start, energy must be provided to break bonds. This initial energy needed is called the activation energy. This energy can also be shown on an enthalpy profile diagram, as in Figure 5.10.

The initial activation energy needed for exothermic reactions often comes from an initial flame, for example, from a match. Once the activation energy is overcome, the net output of energy provides the activation energy for the reaction to continue. An example of this is the energy from the match flame you need to light the methane in a Bunsen burner. Once the methane is lit, combustion will carry on happening until the methane runs out, as the combustion gives out enough energy to provide activation energy to break bonds.

A balanced equation that has the enthalpy data included is called a thermodynamic equation. For example, the thermodynamic equation for the reaction of carbon (graphite) with excess oxygen is:

$$C(s) + O_2(g) \rightarrow CO_2(g)\ \Delta H^\circ = -393.5\ \text{kJ mol}^{-1}$$

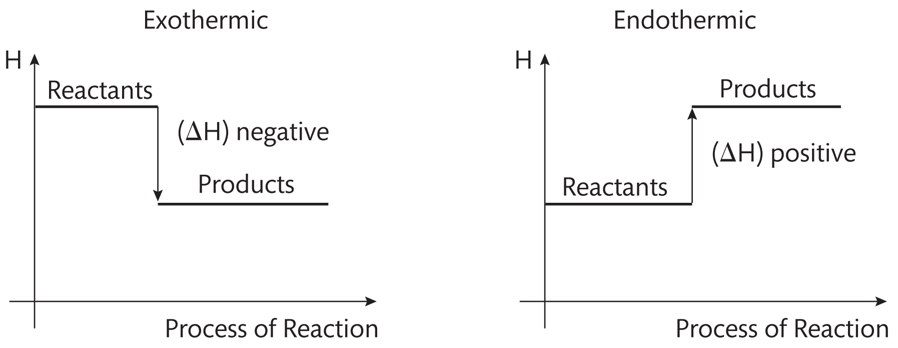
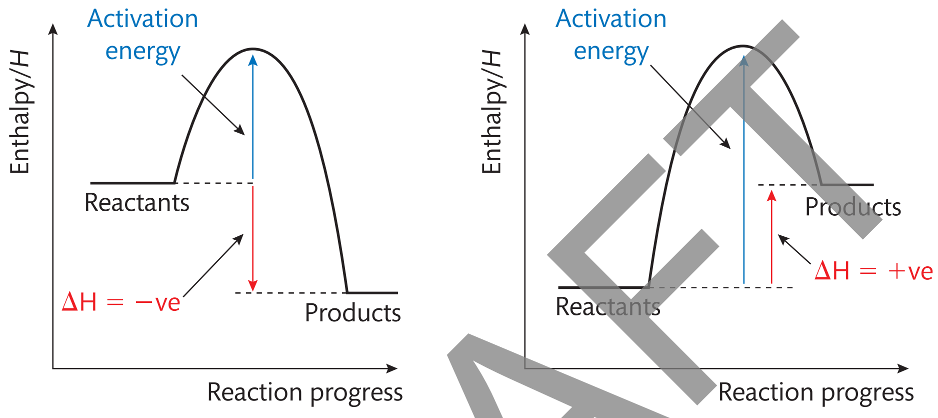


Figure 5.9 Exothermic and endothermic enthalpy changes



Exothermic reaction

Endothermic reaction

Figure 5.10 Activation energy for exothermic and endothermic reactions

When writing a thermodynamic equation, you may need to use fractions to balance the equation, to ensure there is only one mole of reactant.

Enthalpy change of formation is the energy change that takes place when one mole of a compound is formed from its constituent elements in their standard state under standard conditions. It has the symbol $\Delta_f H^\circ$. The f stands for formation.

Enthalpy change of combustion is the energy change when one mole of a substance is completely burned in oxygen. It has the symbol $\Delta_c H^\circ$. The c stands for combustion. So the thermodynamic equation for the combustion of ethane is

$$C_2H_6(g) + 3\tfrac{1}{2}O_2(g) \rightarrow$$
$$2CO_2(g) + 3H_2O(l)\ \Delta_c H^\circ = -1500\ \text{kJ mol}^{-1}$$

Hydration enthalpy is the measure of the energy released when attractions form between positive or negative ions and water molecules. There are often only loose attractions between the positive ions and the slightly negative oxygen atoms in the water molecules. There may also be dative covalent bonds. Hydrogen bonds form between the lone pairs of electrons on the negative ions and the slightly positive hydrogen atoms in the water.

The size of the hydration enthalpy depends on the attraction between the ions and the water molecules. The smaller the ion, the stronger the attraction will be and the higher the hydration enthalpy. The higher the charge on the ions, the stronger the attraction will be and so the hydration energy will also increase.

You can find standard values (literature values) for enthalpy changes in data books. You should compare any enthalpy changes you have measured or calculated through experiments to these values. Remember that these will be in kJ mol^{-1} and, unless they state differently, they will be measured under standard conditions.

Measurement of enthalpy changes

It is not possible to measure enthalpy, but you can measure enthalpy change. This can be done by measuring the temperature change on a system.

Calorimetry uses a mathematical relationship to calculate enthalpy change from experimental quantitative data.

The mathematical expression used is:

$$Q = mc\Delta T$$

where Q = the heat exchanged with the surroundings measured in joules.

m = the mass of the substance heated or cooled measured in grams.

c = the **specific heat capacity** of the substance heated or cooled, expressed as $\text{J g}^{-1}\text{ K}^{-1}$.

ΔT = change in temperature, measured in kelvin.

Key term

Specific heat capacity – the energy required to raise the temperature of 1 g of a substance by 1 K.

Worked example

0.5 g of ethanol was used to heat 100 cm^3 of water from 21°C to 45°C.

What is the molar enthalpy change of this reaction?

Specific heat capacity of water is $4.18\text{ Jg}^{-1}\text{ K}^{-1}$

Use $Q = mc\Delta T$

Mass – mass of 1 cm^3 of water has mass of 1 g so 100 cm^3 of water has a mass of 100 g

$c = 4.18$

$\Delta T = 45 - 21$

$= 24$ K as 1°C is equivalent to 1 K

$Q = 100 \times 4.18 \times 24 = 10\,032\text{ J}$

To calculate molar enthalpy change:

Ethanol has relative molecular mass of 46

$n = \frac{m}{M_r} = \frac{0.5}{46} = 0.011$ so 0.011 moles of ethanol used in this experiment

10 032 J released by 0.011 moles of ethanol so $\frac{10032}{0.011} = 912\,000\text{ J} = 912\text{ kJ mol}^{-1}$ of energy released

The reaction is exothermic and so the enthalpy change is negative.

$\Delta_c H = -912\text{ kJ mol}^{-1}$

However, the reported value is -1367 kJ mol^{-1} so the experimental value is not accurate. This is because it is unlikely that the experiment was carried out under strict standard conditions and incomplete combustion could have occurred. However, the biggest source of error would be the heat lost to the surroundings.

PAUSE POINT Define the following enthalpy changes: combustion, formation, hydration.

Hint You can use words and equations.

Extend Draw a step by step guide showing how you would calculate the molar enthalpy change of a burning fuel.

Further reading and resources

www.rsc.org The website of the Royal Society of Chemists.
www.sciencebuddies.org A website giving hands on science projects.
www.virtlab.com A series of hands on experiments and demonstrations in chemistry.
www.iupac.org The website of the International Union of Pure and Applied Chemistry.

Getting started

This unit builds on Unit 1, where you learned about cells and tissues. It is very important that scientists, science technicians, medical personnel, teachers, nursery nurses and others have an understanding of the human body and its organ systems. This knowledge can be used in many ways, including to assess and treat illnesses, and to help prevent disease. It is essential that you understand the body's structure and function in terms of organs and organ systems, and how they all work together to maintain a steady internal environment known as **homeostasis**. List any organs and organ systems you can think of, and briefly note the importance of each system. When you have completed this unit, see if you can add more to your list.

B Organs and systems

The cardiovascular system

In this section you will learn about the structure and function of the heart, the blood vessels and the events of the **cardiac cycle**. You will also investigate the effect of caffeine on the heart rate in *Daphnia*, more commonly known as water fleas.

Many cells in a multicellular organism are not in direct contact with their surroundings. This means that the organism cannot rely on diffusion alone to supply the cells of all its organs with the nutrients and oxygen it needs to survive. The cardiovascular system, also known as the circulatory system, allows blood to circulate around the body to transport and supply the essential nutrients that multicellular organisms need in order to maintain homeostasis and to survive.

The cardiovascular system consists of the heart, blood vessels and blood (see Figure 5.11). It is responsible for not only transporting nutrients and oxygen but also hormones and cellular waste throughout the body.

Blood is a tissue and it always flows in blood vessels. The components that make up human blood are **erythrocytes**, **leukocytes**, **thrombocytes** and plasma.

Key terms

Cardiac cycle – a complete heartbeat from the generation of the beat to the beginning of the next beat.

Erythrocytes – red blood cells, containing haemoglobin, that transport oxygen around the body.

Leukocytes – white blood cells, of which there are many different types.

Thrombocytes (platelets) – component of blood involved in blood clotting.

Blood is responsible for transporting:

- oxygen from the lungs to body cells for aerobic respiration
- carbon dioxide from respiring cells back to the lungs to be exhaled
- nutrients from the intestines to body cells
- urea (waste product) from the liver to the kidney to be removed
- hormones from the endocrine glands to their target cells
- heat from respiring tissue to organs to maintain body temperature or to the skin to be lost.

Blood is also responsible for regulating:

- body temperature
- pH of body tissues
- volume of fluid in circulation.

Blood also protects us because it contains:

- platelets that cause clotting, to prevent bleeding and entry of pathogens
- leukocytes that defend us against infection.

Link

Go to *Unit 20: Biomedical Science*, Section A, to find more information about the structure and function of erythrocytes, leukocytes and thrombocytes.

Characteristic features of blood vessels

Blood is always in blood vessels, either arteries, arterioles, capillaries, venules or veins. The vessels form a closed transport system which starts and finishes at the heart. We refer to the mammalian circulation system as double circulation because we have two separate circulations. The blood flows from the heart to body tissues and back to the heart. This is known as **systemic circulation**. In a separate

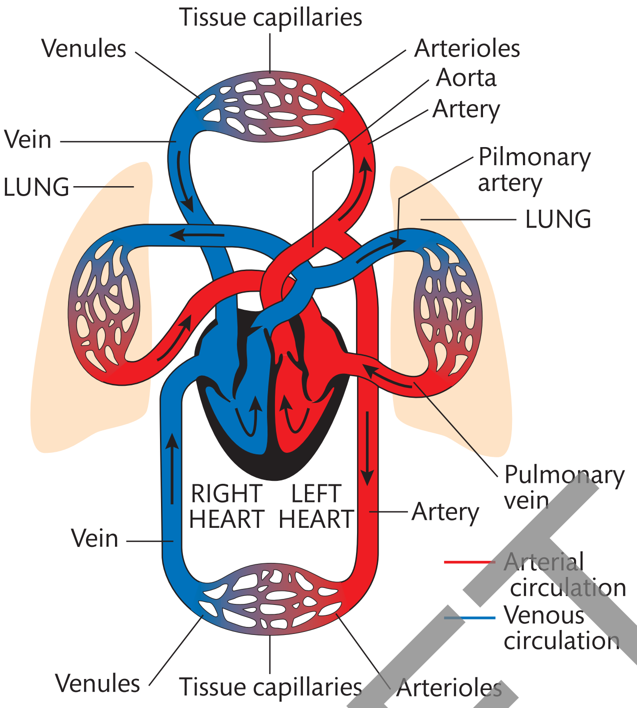


Figure 5.11 The cardiovascular system structure: heart, blood and blood vessels

circulation known as **pulmonary circulation**, blood flows from the heart to the lungs to expel carbon dioxide and take in oxygen and then it returns back to the heart.

Key terms

Systemic circulation – parts of the circulatory system concerned with the transport of oxygen to, and carbon dioxide from, the heart to body cells.

Pulmonary circulation – parts of the circulatory system concerned with the transport of oxygen to, and carbon dioxide from, the heart to lungs.

Arteries and arterioles

Arteries carry blood away from the heart. The blood leaves the ventricles of the heart and enters thick elastic arteries under high pressure. The artery lumen is small and arteries have thick walls that contain collagen, a fibrous protein, to help maintain the shape and volume of the arteries under pressure. Artery walls also contain elastic tissue to enable them to continuously expand and recoil to keep the blood under pressure. This expansion and recoiling is what you will feel as your pulse where an artery passes near the surface of the skin. Arteries also contain smooth muscle that contracts, enabling the artery lumen to narrow if needed. Arteries are lined with smooth endothelium tissue, reducing the friction as the blood flows through the lumen. As blood flows further from the heart, these large elastic arteries become smaller muscular arteries which carry the blood to the organs in the body. They contain less elastic tissue and more smooth muscle than larger elastic arteries. Arteries then divide into smaller arterioles. These vessels contain smooth muscle cells wrapped around the endothelium. These eventually become capillaries.

Capillaries

Capillaries allow the exchange of materials between blood and the body's cells via tissue fluid. These are tiny vessels with very thin walls consisting of only one layer of endothelium cells. This thin wall reduces the diffusion distance for the materials being exchanged. The lumen of a capillary is very narrow, with a diameter that is the same size as a red blood cell. The small diameter allows only one erythrocyte through at one time. This ensures that the red blood cell has to squeeze through the capillaries which help it release the oxygen. Capillaries spread throughout tissues, forming capillary networks, and it is here where the plasma of blood, rich with nutrients and oxygen, is forced out through small gaps in the capillary walls. This fluid that is forced out is known as tissue fluid and it also carries away the waste from cellular activity. Capillaries link arterioles to venules.

Venules and veins

Capillaries become slightly larger and form venules, which also have small diameters. Venules join to make veins. Veins have a large lumen and their walls are thinner than arteries. Veins have thinner layers of collagen, smooth muscle and elastic tissue as they do not need to constrict and recoil. Veins have valves to help prevent the back flow of blood as it makes its way back to the heart. The action of the surrounding skeletal muscle can flatten veins. This also helps force the blood back to the heart.

Table 5.8 states the role of the major blood vessels found in the body.

Table 5.8 Roles of the major blood vessels

Blood vessel	Artery or vein	Role
Vena cava	Vein	To deliver deoxygenated blood from the body into the right atria of the heart.
Pulmonary vein	Vein	To deliver oxygenated blood from the lungs into the left atria of the heart.
Pulmonary artery	Artery	To transport deoxygenated blood away from right ventricle in the heart to the lungs to collect oxygen.
Aorta	Artery	To transport oxygenated blood away from left ventricle in the heart to the rest of the body.
Coronary Artery	Artery	To supply the cardiac muscle with its own supply of oxygen.

PAUSE POINT Explain the structure and function of blood and the vessels it travels in.

Hint Close the book and list the components of blood and all the blood vessels.

Extend Think about the differences between the components and the vessels and why their structure is important for their function.

Structure and function of the heart

Your heart is about the size of your fist when it is clenched and it has a mass of about 300 g. It is in the thoracic cavity between the lungs and behind the sternum, enclosed in a fibrous bag made from inelastic connective tissue called the **pericardium**. The heart is a muscular double pump divided into two halves, each of which contain two chambers.

The wall of the heart is made from mainly **myocardium**, which consists of cardiac muscle. Cardiac muscle contracts to make your heart beat. The coronary arteries seen in Figure 5.12 lie on the surface of the heart. They carry oxygenated blood to the heart muscle itself.

Key terms

Pericardium – a fibrous membrane that surrounds and protects the heart.

Myocardium – the middle and thickest layer of the heart wall, composed of cardiac muscle.

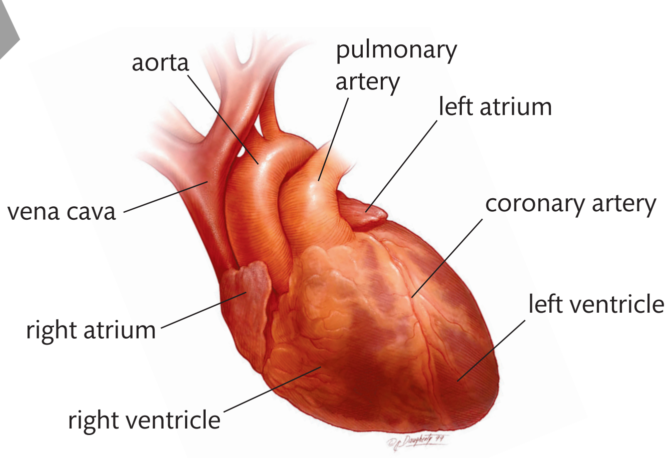


Figure 5.12 External view of the heart

The heart is divided into four chambers. The two upper chambers are called the **atria**. They have thin walls and they sit above the two lower chambers called **ventricles**. Ventricles are thicker walled chambers. Deoxygenated blood flows into the right atrium from the **vena cava** (a main vein); at the same time the left atrium receives oxygenated blood from the lungs via the pulmonary vein. From here the blood flows into the lower ventricles through **atrioventricular valves**. The atrioventricular valve between the left atria and left ventricle is known as the bicuspid valve. The tricuspid is the valve between the right atrium and right ventricle. These valves are thin flaps of tissue attached to the ventricles via tendinous chords to stop the valves from turning inside out. When the ventricles are full of blood and ready to contract, the valves close to stop the back flow of blood into the atrium.

The heart is divided into the right and left side and the ventricles are separated by a wall of muscle called the **septum**. This stops the oxygenated and deoxygenated blood coming into contact with each other.

When the ventricles contract, deoxygenated blood flows upwards out of the right ventricle into the pulmonary artery, where the blood is transported to the lungs to collect oxygen.

Oxygenated blood leaving the left ventricle flows into the aorta. This is a major artery that carries the oxygenated blood to a number of arteries to supply the body cells with oxygen. At the base of both of these arteries there are semi-lunar valves that prevent back flow of blood into the ventricle when they relax.

Key terms

Atria – two top chambers of the heart.

Ventricle – two bottom chambers of the heart.

Vena cava – a large vein carrying deoxygenated blood into the heart.

Atrioventricular valve – the structure found between the atrial and ventricular chambers of the heart to prevent back flow.

Septum – the dividing wall between the right and left sides of the heart.

PAUSE POINT Label the structure of the heart.

Hint Use the internet to find an unlabelled diagram of the heart, and see what you can label.

Extend Draw a flow diagram of the cardiac cycle. Include the functions of all the structures.

The heart muscle is described as myogenic, because it can initiate its own contraction and contracts and relaxes rhythmically, even without stimulation from the nervous system. At the top of the right atrium there is a patch of tissue, commonly known as a pacemaker, called the **sinoatrial node (SAN)**. It is this node that generates the electrical activity and initiates a wave of excitation at regular intervals.

The wave of excitation quickly spreads over the walls of the atrium. As it spreads along the muscle tissue membrane, it causes the muscle cells in both atria to contract simultaneously. This is known as atrial systole. The excitation cannot, however, spread directly to the ventricle because there is non-conducting tissue in the base of the atria. The wave of excitation is picked up by the **atrioventricular node (AVN)** located in the top of the septum separating the two ventricles. The wave of excitation is delayed here, which allows the atria time to complete its contraction and for the blood to flow through the atrioventricular valves into the ventricle before they start to contract.

The AV node stimulates the **bundle of His**, a bundle of conducting tissue made from **Purkinje fibres**. These fibres line the interventricular septum and carry the excitation from the atria to the apex (bottom of the septum in the ventricle). At the apex, the Purkinje fibres spread through the walls of the ventricles and, as the excitation spreads upwards through the muscle tissue, the ventricles contract simultaneously. The ventricles contract from the base upwards and this pushes the blood up into the arteries. This is known as ventricular systole. When the ventricles relax, semi-lunar valves at the base of the aorta and pulmonary artery prevent any blood flowing back into the heart.

Key terms

Sinoatrial node (SAN) – a patch of tissue found in the right atrium that generates the electrical activity and initiates a wave of excitation at regular intervals.

Atrioventricular node (AVN) – a patch of tissue located in the top of the septum that picks up the wave of excitation from the atria.

Bundle of His – a collection of heart muscle cells specialised for electrical conduction.

Purkinje fibres – specialised conducting fibres found in the heart.

The cardiac cycle

One whole cardiac cycle takes about 0.8 seconds.

1. Both atria relax and fill with blood from the pulmonary vein and vena cava. This is atrial diastole.
2. The atria contract and force the atrioventricular (AV) valves open. Blood flows into the ventricles and they fill up; this is ventricular diastole.
3. The AV valves close when the pressure in the ventricles rises above the pressure in the atria to prevent the backflow of blood into the atria.
4. The ventricle walls contract and increase pressure in the ventricles. This forces the semi-lunar valves to open and the blood flows into the pulmonary artery and aorta.
5. When the pressure in the aorta and pulmonary artery rises, the semi-lunar valves close to prevent backflow of blood into the ventricles.

Assessment practice 5.5

1. Describe the cardiac cycle. Ensure that you mention the action of the valves.
2. Explain why the sinoatrial node is commonly referred to as a pacemaker.
3. Explain why the ventricles contract from the apex upwards.

Electrocardiograms

An electrocardiogram (ECG) can be used to monitor the electrical activity of the heart. Some of the electrical activity that is generated by the heart spreads through tissue surrounding the heart and to the skin. During an ECG a number of sensors are attached to the skin, they pick up this electrical excitation and convert it into a trace, as shown in Figure 5.13, that can be interpreted by a medical professional.

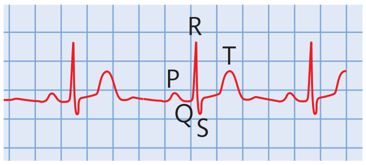


Figure 5.13 Electrocardiogram trace showing a normal ECG

A healthy trace consists of a series of waves and has a particular shape. The waves are labelled P, Q, R, S and T. These waves indicate different information:

- Wave P shows excitation of the atria, when they begin to contract and therefore represents atrial systole.
- Wave QRS indicates excitation of the ventricles, when they begin to contract and therefore represents ventricular systole.
- Wave T shows diastole, when the heart chambers are relaxing.

The shape of an ECG can indicate that there is a problem with the heart muscle. It can detect various arrhythmias:

- Tachycardia is a condition where the heart rate is very fast and over 100 beats per minute. Here the P waves are evenly spread but are closer together than they should be.
- Bradycardia is a condition where the heart rate is very slow and under 60 beats per minute. Here the P waves are evenly spread but further away from each other than they should be.
- Ventricular fibrillation is a very serious condition where the contraction of ventricles is not controlled or coordinated. The ventricles fibrillate/quiver and the heart pumps little or no blood, it can cause cardiac arrest. Here there are no identifiable P, QRS or T waves on the ECG, and the heart rate is very fast.
- Sinus arrhythmia is a normal variation in the beating of your heart. It occurs when your heart rate cycles with your breathing. When you breathe in, your heart rate speeds up and when you breathe out, your heart rate slows back down. Here there is a normal P wave but they are not evenly spread.
- Flat line is where the ECG shows no electrical activity of the heart.

Assessment practice 5.6

1. Describe the shape of an electrocardiogram.
2. Explain why the QRS peak is larger than the P wave.

The effect of caffeine on heart rate

Many different drugs can have an effect on the rate of your heart beat. Caffeine makes your heart 'beat' faster by increasing the electrical activity of the SAN. It also affects the ventricles, leading to an increase in the rate of contraction and relaxation of each heartbeat. As well as beating faster, a larger volume of blood can be pumped out every time the heart beats. Two or three cups of strong coffee or tea contain enough caffeine to increase human heart rate by 5–20 beats/min.

Investigation 5.1

Investigating the effect of caffeine on heart rate in *Daphnia*

Water fleas (*Daphnia*) have a small heart that is very easy to see under a low power light microscope. The heart rate of a daphnia can be up to 300 beats per minute and can be monitored and counted in different conditions. We are going to investigate the effect of caffeine on the heart rate by changing the concentration of caffeine and adding it to the water surrounding the daphnia.

For each step in the investigation, it is important that you understand the purpose of it, and what you need to pay particular attention to, in order that your results are as accurate as possible.

Steps in the investigation	Pay particular attention to . . .	Think about this . . .
1. Take some cotton wool, and place it in the middle of a Petri dish.		
2. Use a pipette to transfer a large daphnia onto the cotton wool fibres.		
3. Add pond water immediately to the Petri dish until the daphnia is just covered by the water.	You should have a good ethical attitude towards the daphnia. Although they are simple organisms that may not 'suffer' in the same way as higher animals, they still deserve respect.	
4. Place the petri dish on the stage of a microscope and observe under low power. The beating heart is located on the dorsal side just above the gut and in front of the brood pouch.		
5. Use a stopwatch to time 30 seconds, and count the number of heart beats in several periods of 30 seconds.	Make sure that you are counting the heart beats, and not the flapping of the gills or movements of the gut. The heart must be observed with transmitted light if it is to be properly visible.	Think about how you might accurately measure the rapid heart rate. For example, to count the beats, make dots on a piece of paper. Once the timer has stopped, count the dots and express heart rate as number of beats per minute (times your dots by 2).
6. Add one drop of water containing caffeine at a concentration of 100 mg/L to 5 cm^3 of pond water in a beaker. Mix well. Draw the pond water off the daphnia with a pipette and replace it with 2 or 3 cm^3 of the water containing caffeine. Record the rate of heart beat again.	Make sure that you measure the volumes accurately, as you will have to repeat this step, and you must keep all the non-variable aspects the same throughout. Scientific investigations produce more accurate results if they are carried out a number of times.	Think about how you might accurately measure heart rate.
7. Repeat this investigation with different concentrations of caffeine, e.g. 10, 1, 0.1, 0.01 and 0.001 mg/L.	Scientific investigations produce more accurate and valid results if they use a large range of data.	In order to find the effect of changing one 'variable' (condition that you can change), you must keep all the other conditions the same in each test you carry out. This is why it is crucial to measure out the volumes accurately each time you repeat the experiment.

Steps in the investigation	Pay particular attention to . . .	Think about this . . .
8. Record your results in an appropriate way and write a report on your investigation.	Your report should inform a reader about how you carried out the investigation, and what you did to ensure accurate results. You should present your results in a way that shows your findings as clearly as possible.	Consider all the ways that you might present your results, e.g. as a table, a graph or a chart. You should aim to make your findings as easy to understand as possible.
9. At the end of the investigation, return the daphnia to the stock culture.	You should have a good ethical attitude towards the daphnia. Although they are simple organisms that may not 'suffer' in the same way as higher animals, they still deserve respect. Animals should be returned to the holding tank after being examined.	

Factors that increase the risk of cardiovascular disease (CVD)

Cardiovascular diseases are diseases that affect the heart and circulatory system. Common diseases are atherosclerosis, coronary heart disease (CHD) and stroke. Factors that increase the risk of CVD are:

- **genetics:** if you have a family history of CVD, then this increases your chance of also suffering CVD
- **age:** as you get older, the risk of CVD increases
- **gender:** statistics indicate that men are at a higher risk of dying of CVD under the age of 50 than women
- **diet:** consuming a high level of saturated fat, high intake of salt, and limited healthy fats and vitamins increases the risk of CVD
- **high blood pressure:** suffering with high blood pressure can increase your risk of a CVD
- **smoking:** smoking cigarettes will increase the risk of suffering with a CVD
- **inactivity :** lack of physical activity will increase risk of suffering with a CVD.

Case study

Coronary heart disease

Dr Amy Bucknell regularly treats people with CVD, in particular people who suffer with coronary heart disease. This is when the blood supply to the heart becomes restricted as a result of the hardening and narrowing of the coronary arteries. One of the common medicines she prescribes are statins. They are usually in tablet form and taken once a day. In most cases they are continued for life. Statins help lower the level of low-density lipoprotein (LDL) cholesterol in the blood. Statins can't cure CVD conditions, but benefits of taking them are that they help prevent the patient's condition from getting worse and statistics tells us that 1 in 50 people taking statins for 5 years will avoid a major heart attack or stroke. One of the important parts of Dr Bucknell's job is to ensure that she discusses the risks of taking statins. Common side effects that affect 1 in 10 patients are nosebleeds, sore throat, headaches, sickness, constipation, increased blood sugar levels and a risk of diabetes. Dr Bucknell also explains that there are further side effects that affect 1 in 100 patients, and these are more severe, such as memory problems, inflammation of the liver and pancreas.

Check your knowledge

1 What may cause arteries to become narrow and restricted?

2 Why are statins an appropriate form of medicine for coronary heart disease?

3 Compare and contrast the benefits and the risks of taking statins.

Benefits and risks of treatments for CVD

Antihypertensives

Antihypertensives are a class of drugs that are used to treat hypertension. Hypertension is commonly referred to as high blood pressure. There are many different types of antihypertensives, which lower blood pressure in different ways. The most widely used drugs are thiazide diuretics, calcium channel blockers, and beta blockers. Table 5.9 lists the advantages and disadvantages of common antihypertensive medication.

Table 5.9 Antihypertensive medication advantages and disadvantages

Antihypertensive	Advantage	Disadvantage
Thiazide diuretics	Lower blood pressure and they are an option for those unable to take or tolerate calcium channel blockers.	They are not suitable during pregnancy and can raise potassium and blood sugar levels so regular blood and urine tests required. They can also cause impotence.
Calcium channel blockers	Lower blood pressure and they are effective in black people and those aged 55 or over.	Not suitable for people with a history of heart disease, liver disease or circulation problems and they can have side effects such as flushed face, headaches, swollen ankles, dizziness, tiredness and skin rashes.
Beta blockers	Lower blood pressure and are an option for people who do not respond to other medication. They are safer to use during pregnancy than other medications.	Cause many side effects such as tiredness, cold hands and feet, slow heartbeat, diarrhoea and nausea. Can also cause sleep disturbances, nightmares and impotence.

Transplantation and immunosuppressants

A heart transplant is a major operation that comes with many risks and complications, but it may be the final option to ensure survival of the patient. These risks include the following.

- **Rejection of the donor heart:** this occurs when the immune system mistakes the new heart as foreign and attacks it. It can happen immediately after surgery or even years after.
- **Infection:** immunosuppressant drugs, taken to reduce the chances of rejection of the donor organ, weaken the immune system and make the patient more vulnerable to infections such as bacterial infections, for example pneumonia, and fungal infections.
- **Failure to pump properly:** the donated heart may not work as it may not start beating, or it may stop beating soon after surgery.
- **Narrowing of the arteries that are connected to the heart:** this is potentially serious as it can restrict the supply of blood to the heart. This could potentially trigger a heart attack.

After a patient has undergone a heart transplant, they need to take immunosuppressants. These can also cause:

- kidney problems
- high blood pressure
- diabetes
- a higher risk of cancer.

Ventilation and gas exchange in the lungs

In this section, you will learn about the structure and function of the lungs and the **ventilation** system. You will also investigate the effect of exercise on the ventilation system using a spirometer to collect data. Respiration is the release of energy from organic molecules found in food; in order for this to occur oxygen is needed.

Key term

Ventilation – the exchange of air between the lungs and the surroundings.

Structure of the human lung and ventilation system

The ventilation system is also known as the respiratory system (see Figure 5.14). It allows oxygen to move into the body and carbon dioxide to be removed from the body as waste. The respiratory system consists of the:

- trachea
- bronchi
- bronchioles
- alveoli
- intercostal muscles
- diaphragm.

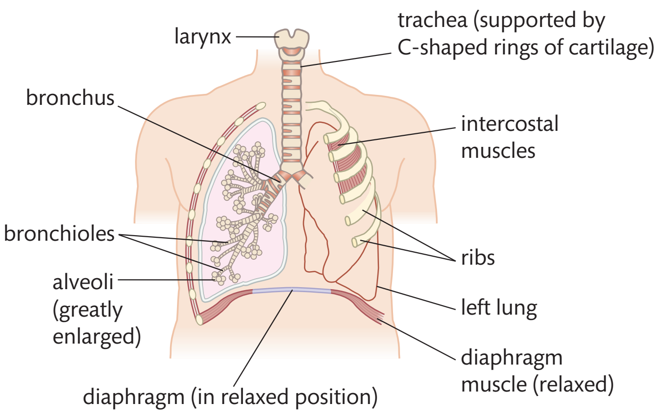


Figure 5.14 The ventilation system structure

Trachea and bronchi

Air is inhaled through the nose and mouth and travels down the trachea, bronchi and bronchioles. These are airways that allow passage of air into the lungs and out again. The trachea and bronchi are similar in structure, they are just different in size. They have thick walls lined with several layers of tissue, and both are supported by walls of cartilage. The trachea has incomplete rings of cartilage, in the shape of a 'C'. The C-shaped rings of cartilage are also present in the bronchi, but they are less regular. Cartilage holds the airways open and prevents them collapsing when there is low pressure. The C-shaped rings of the trachea allow the oesophagus to sit and function alongside it. On the inside of the cartilage there are different layers of tissue.

- Glandular tissue secretes mucus to trap any pathogens and reduce the risk of infection. Further mucus is produced by the goblet cells of the epithelium tissue that lines the airways.
- Smooth muscle can contract to restrict airflow onto the lungs.
- Elastic fibres recoil as the smooth muscle relaxes, helping the airway to widen.

The airways are lined with epithelium tissue, which consists of ciliated epithelium cells and goblet cells. Ciliated epithelium move in a synchronised pattern to waft mucus up and out of the airways.

Link

Go to *Unit 1: Principles and Applications of Science 1* to find more information about tissue structure and function.

Bronchioles

Air continues its journey and travels through the bronchioles. These are much narrower than the bronchi. Bronchiole walls are made mostly of smooth muscle and elastic fibres. Larger bronchioles consist of some cartilage and the smaller bronchioles have no cartilage. The smallest bronchioles have clusters of alveoli at the end.

Alveoli

Finally, when the air reaches tiny sacs called alveoli, gas exchange takes place between the air in the alveoli and the blood in the capillaries that surround the alveoli. Gases pass both ways across the walls of the alveoli to ensure efficient provision of oxygen for respiration, whilst also ensuring that carbon dioxide is removed for efficient production of adenosine triphosphate (ATP) during cellular respiration. The structure of the lungs and the alveoli ensures effective and efficient gas exchange through the following features.

- **Large surface:** there are millions of alveoli, providing more space for the gas molecules to pass through.
- **Short diffusion pathway:** the walls of the alveoli are made of squamous epithelial cells and are only one cell thick. This reduces the distance these molecules have to diffuse through.
- **Capillary network:** each alveolus is close to a capillary which has a well which is also one cell thick to give a short diffusion pathway to molecules.
- **Diffusion gradients:** oxygen diffuses from a high concentration inside the alveoli down a **concentration gradient** to an area of lower concentration in the blood capillary, where it joins with haemoglobin in the erythrocytes. Carbon dioxide diffuses from a high concentration inside the blood capillary down a concentration gradient to an area of lower concentration in the alveoli. If it stayed in the blood it could be toxic because it lowers the pH.
- **Moisture:** a layer of moisture lines the alveoli. Gases can only diffuse across the membrane if dissolved. Therefore this moisture allows gases to dissolve in order to cross.
- **Surfactant:** this is a chemical produced by the lungs to stop the alveoli from collapsing by reducing the surface tension of water.

Key term

Concentration gradient – the difference in the concentration of a substance between two regions.

Ventilation of the lungs

Breathing in is called inspiration. The following steps explain what the body does in order to breathe in oxygen.

1. The intercostal muscles between the ribs contract and raise the rib cage up and out.
2. The diaphragm muscle contracts and the diaphragm flattens and moves down.

3 Both these actions increase the volume in the thoracic cavity.
4 This reduces the air pressure inside the thoracic cavity.
5 As a result, air moves down the trachea, bronchi, bronchioles and into the alveoli from the higher atmospheric air pressure to the lower air pressure of the thoracic cavity.
6 Oxygen diffuses through the alveolar membrane into the blood capillaries and carbon dioxide diffuses from the blood capillaries across the alveolar membrane into the alveoli.

Breathing out is called expiration. This is done in order to push air with a high concentration of carbon dioxide out of the body.

1 External intercostal muscles relax and the rib cage moves down and inwards.
2 The muscles of the diaphragm relax and the diaphragm moves up in a dome shape.
3 Both these actions reduce the volume in the thoracic cavity.
4 This increases the air pressure inside the thoracic cavity.
5 As a result, air is pushed out.
6 In the case of expiration during exercise, the internal intercostal muscles contract to reduce the volume of the thoracic cavity further so that a larger volume of air can be breathed out to get rid of the extra carbon dioxide (CO_2) that has been produced by a higher rate of respiration in muscle tissue.

Role of pleural membranes

The role of the pleural membranes is to protect the lungs because lung tissue is delicate and can be easily damaged. The pleural membranes enclose a fluid-filled space surrounding the lungs which provides lubrication. Our lungs are constantly expanding and contracting so the pleural membranes and fluid enable the lungs to move easily, minimising friction from other organs.

PAUSE POINT Explain the structure and function of the lungs.

Hint Breathe in deeply and think about the passage of air.

Extend How are the alveoli specialised for efficient and effective gas exchange?

Spirometer readings of lung volumes

One way to investigate pulmonary ventilation (breathing) is by using a spirometer. A spirometer consists of a chamber filled with medical grade oxygen that floats on a tank of water. A disposable mouthpiece is connected to a tube. This is connected to the tank, and the patient breathes in and out (see Figure 5.15). Breathing in removes oxygen from the chamber so it moves down, while breathing out pushes carbon dioxide in so the chamber moves up. The movement of the chamber up and down is recorded using a datalogger and produces a trace.

Soda lime is attached to the tube to absorb carbon dioxide that is breathed out. This means that the total volume of gas in the spirometer will gradually decrease. The volume of carbon dioxide breathed out is the same as the volume of oxygen breathed in. Therefore, as carbon dioxide is removed, this total decrease equals the volume of oxygen used up by the person breathing in and out. The trace will show a slope and this can be used to measure the volume of oxygen used in a specific period, such as when exercising.

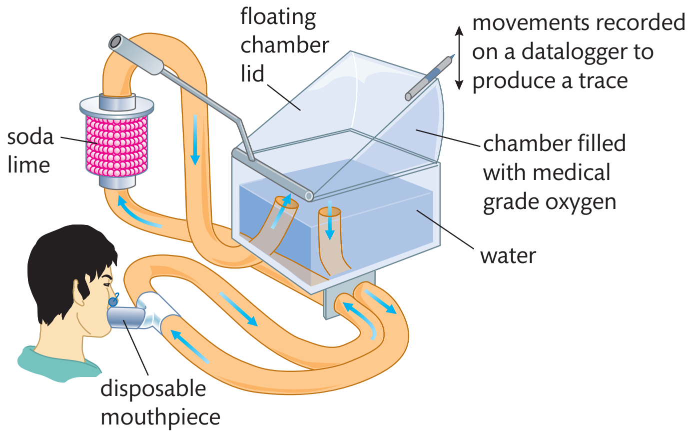


Figure 5.15 Person breathing in and out using a spirometer

Worked example

You can calculate oxygen uptake per minute for a subject at rest.

If a *y*-axis shows a reduction of 0.5 dm^3 between point A and B:

Step 1: Use the *y*-axis for the reduction in oxygen volume which is 0.5 dm^3.

Step 2: Look at the *x*-axis for the time taken: 65 s.

Step 3: So in 65 s, 0.5 dm^3 of oxygen is used up.

Step 4: $0.5/65\ dm^3s^{-1}$.

Step 5: Times it by 60 to work it out per minute $0.5 \times 60/65\ dm^3min^{-1}$.

Step 6: $0.46\ dm^3min^{-1}$.

Types of lung volume

There are different elements of lung volume, as shown in Table 5.10. An example of a spirometer trace showing these elements is shown in Figure 5.16.

Table 5.10 Elements of lung volume

Element	Description
Tidal volume (TV)	The volume of air that is taken in and breathed out during each breath when you are at rest. At rest TV is approximately 0.5 dm^3.
Inspiratory reserve volume (IRV)	The volume of air that can be breathed in when you take a big breath, over and above normal tidal volume. This is approximately 2.5–3 dm^3.
Residual volume	The volume of air that always remains in your lungs even after you have breathed out. This is approximately 1.5 dm^3.
Expiratory reserve volume (ERV)	How much air can be breathed out after you have taken a big breath in, over and above normal tidal volume. This is approximately 1 dm^3.
Vital capacity	The maximum volume of air that can be moved in and out of your lungs in one breath. This varies depending on size, age and sex of the person; it is approximately 5 dm^3.
Total lung capacity	This can be calculated by adding together the person's vital capacity and the residual volume. This will provide you with data to suggest the total lung volume of a person.

The air in the nose, trachea and bronchi is not involved with gas exchange and is known as dead space.

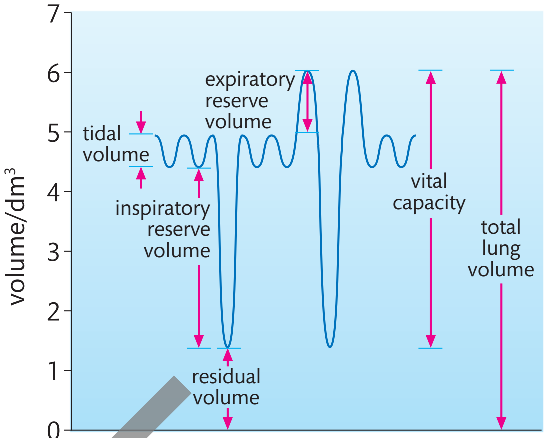


Figure 5.16 A spirometer trace showing tidal volume, residual volume, inspiratory reserve volume, expiratory reserve volume, vital capacity and total lung capacity

Methods used to measure lung function for respiratory conditions

Medical personnel may need to measure the function of people's lungs on occasion to determine if there is an underlying medical condition that is stopping them from functioning properly.

Peak expiratory flow

A peak flow measures the speed of air flowing out of a person when they breathe out as fast and as much as they can. A peak flow reading can indicate how well the lungs are functioning. A peak flow is commonly used to determine how well an asthmatic person's lungs are working. The test is carried out three times and the best reading is recorded.

Forced vital capacity

Forced vital capacity (FVC) is a lung function test that can be measured using a spirometer. It is used to diagnose obstructive lung diseases such as asthma and chronic obstructive pulmonary disease (COPD).

To achieve data for forced vital capacity, the patient breathes out as forcefully and rapidly as possible into a spirometer. The forced vital capacity is the total volume of air exhaled; it is normally equal to the vital capacity. It can help determine the amount of obstruction that a person has in their airways.

The effects of exercise

When you are at rest, air moves in and out of your lungs about 12 times per minute. Each breath renews the air in your lungs and expels carbon dioxide from your body.

When you exercise or become frightened, your breathing becomes quicker and deeper to supply your body with more oxygen and remove more carbon dioxide. This is known as breathing rate.

The frequency at which we take breathes in must increase to meet the demand for oxygen. This therefore increases the **respiratory minute ventilation**, which is the volume of air breathed in or out per minute. If a person were to carry out exercise and then breathe in and out through a spirometer, we would notice a difference in tidal volume.

Key term

Respiratory minute ventilation – volume of air breathed in or out per minute.

Exercise causes tidal volume to increase, to meet the oxygen demands of respiring muscles and to accommodate the exhalation of the increased production of carbon dioxide. During exercise, your body consumes large amounts of oxygen. The harder you exercise, the more oxygen your body consumes.

Assessment practice 5.7

1. Explain how a spirometer works.
2. Explain the data that can be analysed from a spirometer.
3. Use the spirometer trace to determine the oxygen uptake per minute.

Urinary system structure and function

In this section you will learn about the structure and function of the urinary system, which is also known as the renal system. You will learn about the purpose of the urinary system in excretion and **osmoregulation** and how the kidneys are involved in water, electrolyte and acid base balance. The section also covers problems with kidney function and how kidney disease can be treated.

Key term

Osmoregulation – the control of water and salt levels in the body which prevents problems with osmosis.

The function of the urinary system

The urinary system consists of the kidney, ureters, bladder and urethra. Most people have two kidneys. They are positioned each side of the spine at the back of the abdominal cavity just below your waist. The ultrastructure of the kidney will be discussed later in this unit. Figure 5.17 shows the structure of the urinary system.

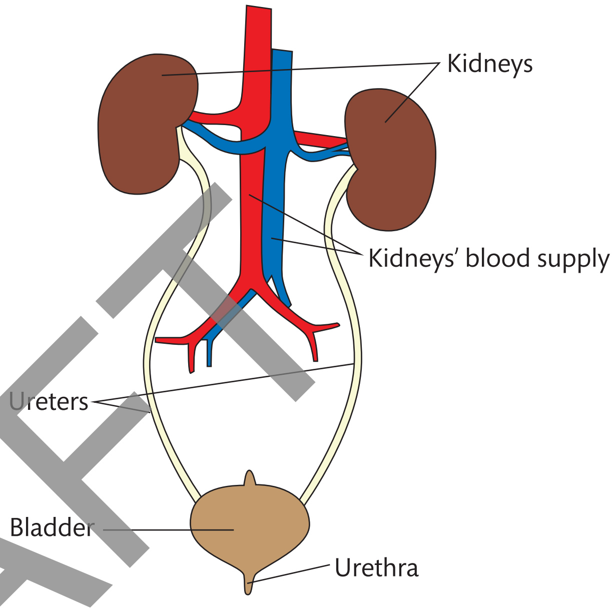


Figure 5.17 Structure of the urinary system

Ureter

The ureters are muscular tubes made of smooth muscle fibres. They transport urine from the kidneys to the bladder.

Bladder

The urinary bladder is a hollow muscular organ that collects urine from the kidneys before disposal by urination. The urine enters the bladder through the ureter and urine leaves the bladder by the urethra.

Renal artery and vein

There are two types of blood vessel attached to the kidneys. The renal arteries deliver an oxygen-rich blood supply to the cells in each kidney. Once the blood has been processed here it leaves the kidney via the renal veins and is transported in the inferior vena cava back to the heart.

The kidney

The kidney filters waste products from the blood before turning them to urine. The kidney consists of three very easily identifiable regions (see Figure 5.18). It is surrounded by a strong capsule and consists of the cortex, medulla and in the centre the renal pelvis which leads to the ureter.

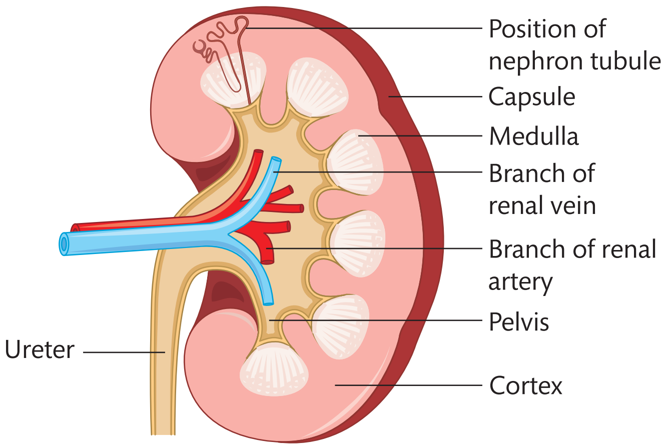


Figure 5.18 The structure of the kidney

The structure and function of a kidney nephron

The nephron is the functional unit of the kidney. Urine is produced here. Nephrons are microscopic structures that make up the bulk of the kidney. Each kidney has approximately one million nephrons, and each is close to many blood capillaries (see Figure 5.19).

Each nephron starts in the cortex of the kidney. Here, the blood capillaries from the renal artery form a knot known as the glomerulus. This sits inside a cup-shaped structure called the Bowman's capsule. Ultrafiltration is the process whereby fluid from the blood is pushed into the Bowman's capsule so that the process of selective reabsorption can take place as the fluid flows along the nephron. Substances that the body needs to conserve are reabsorbed back into the blood capillaries and anything not reabsorbed ends up as urine to be expelled from the body.

The nephron is divided into four parts:

- proximal convoluted tubule
- loop of Henle
- distal convoluted tubule
- collecting duct.

Ultrafiltration and the glomerulus

The glomerulus receives blood from the **afferent arteriole** and the blood leaves through the efferent arteriole. The afferent arterioles are wider than the efferent arterioles; this difference in diameter increases the pressure in the blood capillaries of the glomerulus and pushes fluid out of the capillaries and into the Bowman's capsule where there is a lower pressure (see Figure 5.20). The fluid pushed out of the blood capillaries contains water, amino acids, glucose, urea and inorganic ions for example sodium, chloride and potassium. Blood cells and proteins are left in the capillary as they are too large

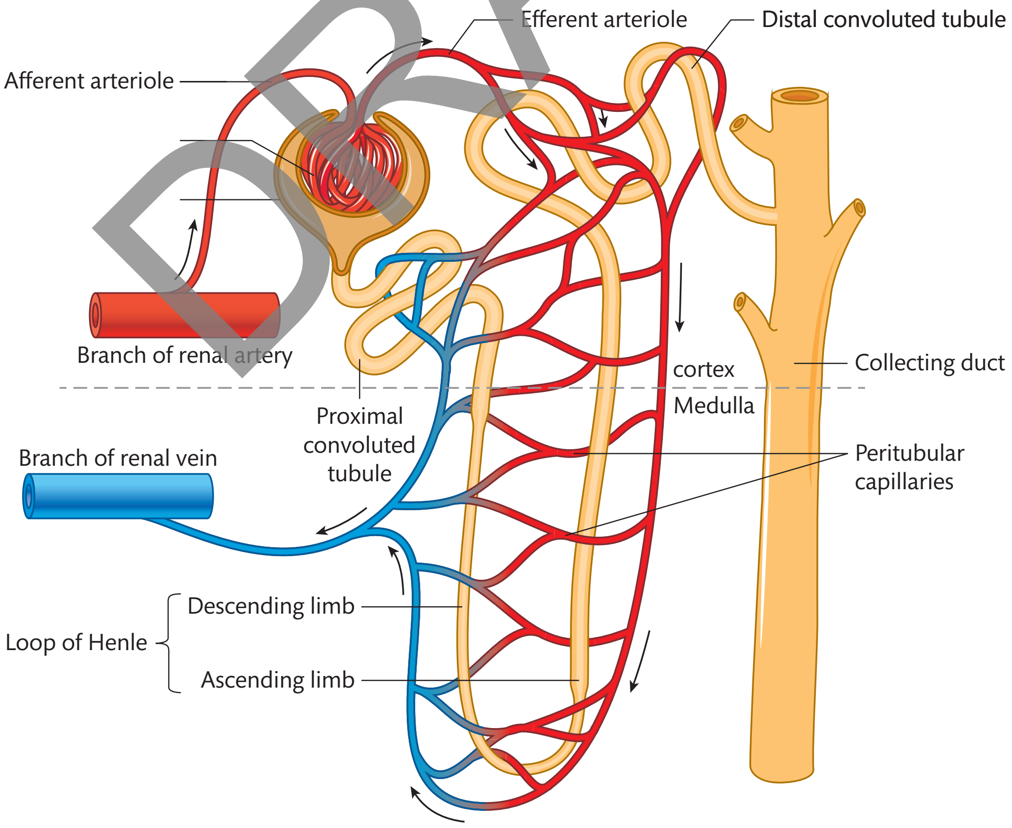


Figure 5.19 A nephron and blood capillaries

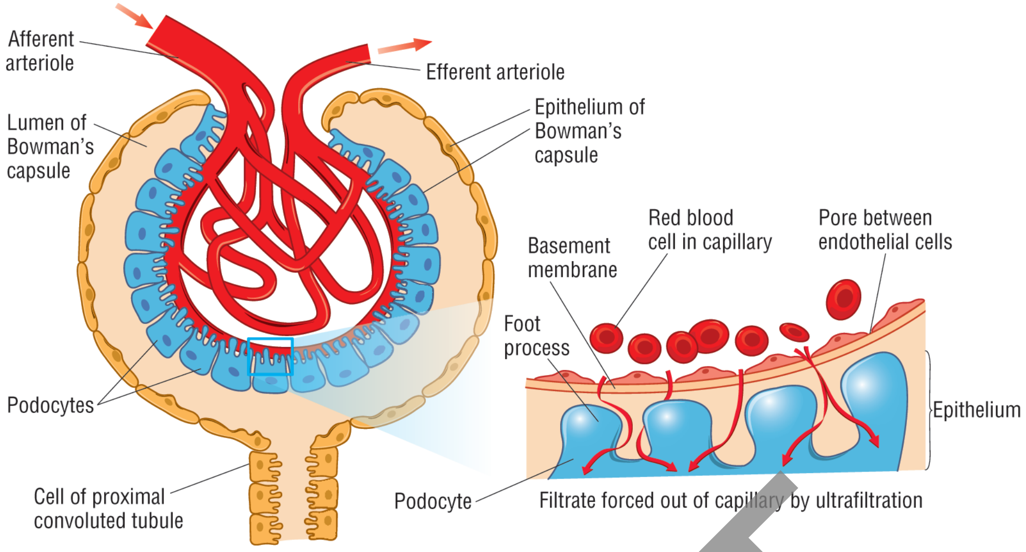


Figure 5.20 Glomerulus and Bowman's capsule

to pass through the small gaps in the capillary wall. This means the **water potential** in the blood capillaries is very low after ultrafiltration. This is important to help reabsorb water later in the process.

Key terms

Afferent arterioles – a group of blood vessels that supply the nephrons in many excretory systems.

Water potential – a measure of the ability of water molecules to move freely in solution.

The fluid must pass through three layers to get into the Bowman's capsule.

- **Endothelium of the blood capillary:** there are small gaps in between the cells that line the blood capillary so that blood plasma and the substances dissolved in it can pass through.
- **Basement membrane of Bowman's capsule:** this consists of a very thin network of collagen fibres and glycoproteins. It acts as a filter to prevent the movement of larger substances from the blood capillary into the Bowman's capsule.
- **Epithelial cells of Bowman's capsule:** these cells are called podocytes and have many finger-like protections called foot processes to ensure there are gaps between the cells. This enables the fluid from the capillary to pass between the cells into the Bowman's capsule.

Selective reabsorption and the nephron

As the fluid moves along the nephron, all the glucose and amino acids, and some salts and water that have been pushed out of the blood capillaries are reabsorbed back into the blood. 85% of the filtrate is reabsorbed in the proximal convoluted tubule. The cells that line the proximal convoluted tubule are specialised for this reabsorption.

The proximal convoluted tubule cell surface membrane that is in contact with the fluid in the tubule is highly folded into microvilli to increase the **surface area**. This increases the space available for reabsorption to take place. The membrane also contains co-transporter proteins that transport proteins and glucose along with sodium ions from the convoluted tubule into the cells lining the tubule by **facilitated diffusion**. The opposite surface of the membrane that is situated next to the blood capillaries is highly folded and contains sodium-potassium pumps. These pump sodium out of the cells lining the convoluted tubule and potassium ions in. This reduces the concentration of sodium in the cells' cytoplasm. This means that sodium ions from the filtrate are transported into the cells, by facilitated diffusion, from the tubule along with glucose and amino acids also from the filtrate. This increases the concentration of glucose and proteins inside the cell so they are able to diffuse from a high concentration inside the cell into the tissue fluid that surrounds the cells and back into the blood capillaries. The reabsorption of salts, glucose and proteins reduces the water potential in the cells that line the tubule and increases the water potential in the tubule. This means water will also move into the cells down a water potential gradient from an area of high water potential in the tubule to an area of low water potential in the cells lining the tubule. From the cells it can be reabsorbed back into the blood capillaries.

Link

Go to *Unit 9: Human Regulation and Reproduction* to find more information about the sodium potassium pump.

Key terms

Surface area – a measure of the total space occupied by the surface.

Facilitated diffusion – the movement of molecules down their concentration gradient, across a membrane, with the help of carrier proteins. Energy is not required.

The loop of Henle

The loop of Henle creates a low water potential in the surrounding medulla tissue, to ensure that water can be reabsorbed from the nephron and back into the blood capillaries. The loop of Henle consists of:

- the descending limb, which descends from the cortex into the medulla
- the ascending limb, which ascends from the medulla into the cortex.

The arrangement of the loop of Henle creates a hairpin counter current. The effect of this arrangement is to increase the efficiency of salt transfer from the ascending limb to the descending limb. This causes a high concentration of salt (sodium chloride) to build up in the surrounding tissue, so that it diffuses out of the loop of Henle into the medulla tissue. This reduces the water potential of the medulla tissue so water moves out of the collecting duct and can therefore be reabsorbed. The steps below show how this is achieved.

Step-by-step: loop of Henle

7 Steps

1 At the base of the ascending tubule, sodium and chloride ions diffuse out of the tubule into the tissue fluid; this reduces the water potential of the surrounding tissue.

2 As fluid moves down the descending limb the water potential inside the tubule becomes lower.

3 As the fluid ascends up the ascending limb towards the cortex the water potential inside the tubule increases.

4 Higher up the tubule sodium and chloride ions are actively transported out into the surrounding tissue fluid. This reduces the water potential of the surrounding tissue.

5 The wall near the top of the ascending limb is impermeable to water, so water can't leave the tubule. This means that the fluid in the ascending limb loses salts but not water.

6 A consequence of the movement of salts from the ascending limb is that water moves out of the descending limb by **osmosis** into the surrounding tissue fluid where the water potential has become lower.

7 Also, sodium and chloride ions diffuse into the descending limb from the surrounding tissue, as the concentration of these ions is higher in the tissue than in the tubule, which decreases the water potential in the tubule.

The water potential in the medulla becomes more negative the deeper the nephron goes into the medulla. The removal of salts from the ascending limb means that the urine at the top of the ascending limb is very dilute so the water in this urine can be reabsorbed in the distal tubules and collecting ducts. The amount of water we reabsorb depends on the body; this is referred to as osmoregulation, the regulation of water.

Key terms

Osmosis – the net movement of water molecules from an area of high water potential to an area of low water potential, down a water potential gradient and across a partially permeable membrane.

Osmoregulation - the control of water and salt levels in the body which prevents problems with **osmosis**.

Distal convoluted tubule and collecting duct

The fluid leaves the ascending loop of Henle and passes along the distal convoluted tubule, where **active transport** is used to adjust the concentration of salts. The fluid, which is still very dilute and contains a high concentration of water, flows from here into the collecting duct. The collecting duct descends into the medulla to the renal pelvis. The fluid in the tubule passes down the collecting duct and, because of the low water potential in the surrounding tissue, water moves out of the collecting duct by osmosis into the surrounding tissue and enters the blood capillaries to be carried away and used in the body. The amount of water that is reabsorbed depends on the permeability of the walls of the collecting duct. This is controlled by antidiuretic hormone (ADH).

Key term

Active transport – the movement of molecules against a concentration gradient, across a membrane.

PAUSE POINT Explain ultrafiltration and selective reabsorption.

Hint Think about the ultrastructure of the kidney and the functional unit.

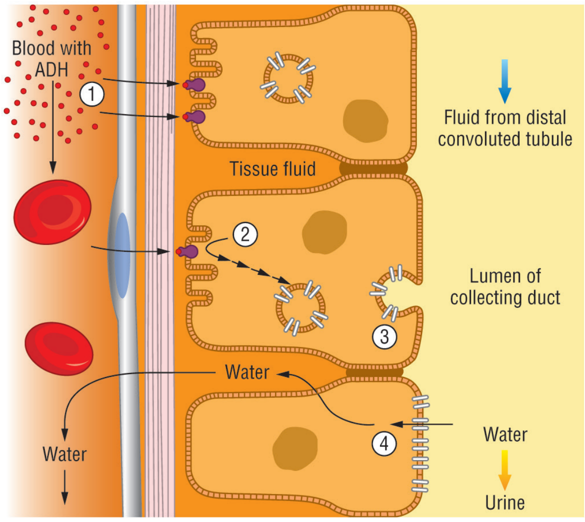
Extend Explain the change in water potential throughout the medulla of the kidney and how this affects reabsorption of substances.

Osmoregulation and the role of anti-diuretic hormone (ADH)

In order to take water in, we drink and eat. We also produce water in respiration. Water is lost from our body in urine, sweat, faeces and water vapour when breathing out. The correct water balance between cells and the fluid that surrounds them must be controlled. **Osmoregulation** is the control of water and salt levels in the body which prevents problems with osmosis. You may notice that sometimes you have very small amounts of concentrated urine but on other days you have large amounts of dilute urine.

Depending on your body's needs, the walls of the collecting duct can be made more or less permeable. For example, if it is a hot day, you need more water, so the walls of the collecting duct are made more permeable so that more water can be reabsorbed into the blood to keep levels of water high. This will mean you will produce a smaller volume of urine. On a cooler day, when you do not need as much water, the walls of the collecting duct are made less permeable so that less water is reabsorbed and you lose more water in your urine.

Antidiuretic hormone (ADH) is manufactured by special cells called neurosecretory cells. These are found in the hypothalamus in the brain. ADH is released into the blood and binds to receptors on the wall of the collecting duct of the nephron. This causes a chain of enzyme-controlled reactions to occur inside the cells. The outcome of these reactions is that water permeable channels called aquaporins are inserted into the collecting duct cell surface membrane (see Figure 5.21). These channels facilitate the movement of water and more channels mean that more water can be reabsorbed, by osmosis, from the collecting duct across the membrane and into the blood capillaries so less water leaves the body in urine. Less ADH in the blood means that there are fewer aquaporins, because the cell surface membrane folds inwards to create vesicles that contain the water permeable channels. This leaves behind fewer channels, so the collecting duct walls are less permeable to water. Less water is reabsorbed and therefore more water leaves the body in the urine.



1. ADH detected by cell surface receptors
2. Enzyme-controlled reactions
3. Vesicles containing water-permeable channels (aquaporins) fuse to membrane
4. More water can be reabsorbed

Figure 5.21 Effect of ADH on the collecting duct wall

Step-by-step: Osmoregulation

8 Steps

1 As blood flows through the hypothalamus in the brain, osmoreceptors in the hypothalamus monitor the water potential of the blood.

2 If the water potential in the blood is low, then the osmoreceptor cells lose water by osmosis and they shrink.

3 This stimulates neurosecretory cells in the hypothalamus to produce ADH. ADH is made in the cell body of these special neurons.

4 ADH flows down the axon of the neurosecretory cell to the terminal bulb in the posterior pituitary gland, where it is stored until it is needed.

5 When the neurosecretory cells are stimulated, they send action potentials down their axons and the ADH is released into the blood capillaries running through the posterior pituitary gland.

6 ADH is transported around the body and acts on the cells in the collecting ducts.

7 When the water potential rises, the osmoreceptors in the hypothalamus detect this change, and less ADH is produced.

8 ADH has a **half-life** of about 20 minutes, so ADH present in the blood is broken down and the collecting ducts will not be stimulated any further.

Key term

Half-life – the time it takes for the concentration to reduce to half of its original value.

Assessment practice 5.8

1 How do neurosecretory cells differ from normal nerve cells?

2 Explain how the kidney controls water potential in the blood.

3 Why is it important that ADH is broken down and does not remain in the blood?

Kidney disease

Medical problems such as kidney failure can occur due to many different reasons. Some of the more common causes are diabetes mellitus, hypertension and infection. If the kidneys fail completely, the body is not able to control the levels of water in the body. It is unable to remove excess water and waste products such as urea, and this can lead very quickly to a fatality.

Case study

Haemodialysis

Lillian has kidney failure and has to undergo haemodialysis at a clinic three times a week. Each session is several hours. Dialysis removes waste and excess fluid and salt from Lillian's blood. Blood from Lillian's vein is passed through a dialysis machine that contains an artificial dialysis membrane. The membrane is partially permeable and allows the exchange of substances between Lillian's blood and dialysis fluid. The dialysis fluid contains the correct concentrations of salts, water, urea and other substances that are required in Lillian's blood plasma. If there are high concentrations of any of these substances in Lillian's blood, then they will diffuse out of the blood across the membrane into the dialysis fluid. If Lillian's blood has low concentrations of any of the substances, then they will diffuse across the membrane from the dialysis fluid into her blood. This helps to achieve the correct concentrations of substances in Lillian's blood.

Check your knowledge

1. What is the function of dialysis fluid?
2. Why must the dialysis membrane be partially permeable?
3. How may this impact on Lillian's everyday life?
4. What are the benefits for Lillian?

Case study

Peritoneal dialysis

Mark is researching peritoneal dialysis as an alternative to Lillian having to attend a clinic three times a week. He has found that peritoneal dialysis uses a membrane in your body called the peritoneal membrane. It requires a surgeon to implant a permanent tube in the abdomen so that the dialysis fluid can be poured into the abdomen space to allow exchange of substances across the peritoneal membrane. After several hours the solution is drained. Lillian would have to perform 4–6 exchanges every day, but it would mean she could do it in the comfort of her own home, when it is suitable for her.

Check your knowledge

1. What are the main differences between haemodialysis and peritoneal dialysis?
2. What are the advantages to Lillian carrying out peritoneal dialysis?
3. What are the disadvantages of changing to peritoneal dialysis?

Kidney transplant

A kidney transplant involves major surgery, where the patient is under anaesthetic and the surgeon implants the donor kidney into the lower abdomen and attaches it to the blood supply and bladder. Patients are given immunosuppressant drugs to help prevent rejection of the foreign organ.

Assessment practice 5.9

1. Explain why haemodialysis fluid must be sterile and at 37°C.
2. Create a table of advantages and disadvantages for dialysis and kidney transplant.

Cell transport mechanisms

For all the processes mentioned in this unit (e.g. reabsorption of glucose, water and gas exchange in the lungs), cell membranes must have a specific structure to enable the movement of substances from one place to another.

Fluid mosaic model

The term **fluid mosaic model**, proposed by Singer and Nicholson, is used to describe the arrangement of biological membranes. Fluid mosaic membranes consist of the following layers.

- The phospholipid bilayer, which forms the basic structure.
- Protein molecules, which are present within the phospholipid bilayer.
- Extrinsic proteins, which are embedded on the surfaces of the membrane.
- Intrinsic proteins, which completely span the bilayer.

Phospholipid bilayer

Phospholipid molecules consist of a phosphate head and two fatty acid tails. The phosphate head is described as **hydrophilic** (water-loving) due to the charges distributed across the molecule while the two fatty acid chains are **hydrophobic** (water-hating).

Key terms

Fluid mosaic model – description of the cell membrane structure, a phospholipid bilayer with proteins floating in it.

Hydrophilic – associates with water molecules easily.

Hydrophobic – repels water.

When a phospholipid molecule is mixed with water the heads stick into the water, while the fatty acid tails stick up out of the water. When phospholipids are completely surrounded by water, they form a bilayer.

All biological membranes are made from phospholipid bilayers, and the hydrophobic layer formed by the fatty acid tails creates a barrier that helps separate the cell contents from the outside. All membranes are permeable to water because it can easily diffuse through the bilayer. It is also very important for survival that cells get a supply of nutrients that they need and that any waste produced is removed from the cell. These molecules usually enter or leave the cell across the membrane.

Methods used to transport molecules through cell membranes

Diffusion

Diffusion is the movement of molecules from an area of high concentration to an area of lower concentration down a concentration gradient. Molecules possess kinetic energy which keeps them moving and they are passively transported across biological membranes. Diffusion is therefore known as a passive process as the molecules only rely on their kinetic energy and a concentration gradient for movement, they do not use energy from the cell.

Lipid-based molecules

Fat-soluble molecules such as steroid hormones can simply pass through the phospholipid membrane because the bilayer consists of fatty acid tails. They diffuse down a concentration gradient through the membrane and into the cell.

Key term

Fat soluble – dissolves in fats.

Small molecules

Oxygen and carbon dioxide molecules are small enough to just pass through the spaces in the phospholipid bilayer and can be transported across phospholipid bilayers by diffusion.

Facilitated diffusion

Larger and charged molecules

Large molecules, such as glucose, and small charged particles, such as sodium **ions,** are not able to pass through the phospholipid bilayer. They need help to cross the membrane. This is known as facilitated diffusion. There are two types of proteins present in the membrane that facilitate diffusion.

- **Channel proteins** form pores in the membrane which are shaped to allow particular molecules/ions, for example sodium and calcium ions, to pass through. Many are 'gated', which means they can be open and closed.
- **Carrier proteins** are shaped for a specific molecule, for example glucose or amino acids. When the molecule binds to the protein, the protein changes shape to allow the molecule to pass across the membrane.

Key term

Ions – particles that carry a positive or negative charge.

Active transport

Carrier proteins in the membrane can act as pumps to carry large and charged molecules across the membrane. The shapes of the proteins are complementary to the molecules they carry, which they transport one way across the membrane. As a molecule moves through the protein, its shape changes. This means that as a molecule exits it cannot enter again, as the protein shape is no longer complementary. These protein pumps use metabolic energy in the form of ATP (adenosine triphosphate) to move molecules across the membrane and they can carry molecules in the opposite direction to the concentration gradient from a low concentration to a high concentration. This process of active transport is much faster than diffusion.

Endocytosis and exocytosis

Sometimes large quantities of materials need to be moved into cells, by **endocytosis**, or out, by **exocytosis**.

Bulk transport requires energy in the form of ATP. The energy is used to move the membrane around to form and move vesicles around the cell. Vesicles are used to carry the bulk material, for example insulin, to be transported and they easily fuse with membranes and can separate from membranes by 'pinching off'.

Key terms

Endocytosis – movement of bulk material into a cell.

Exocytosis – movement of bulk material out of a cell.

Osmosis

Osmosis is diffusion of water molecules only. Water potential is the pressure exerted by water molecules. If there is a higher concentration of water molecules they will exert a higher pressure and therefore have a higher water potential. You must always refer to osmosis as the movement of water molecules from a region of high water potential to a region of low water potential, down a water potential gradient across a partially permeable membrane. Osmosis will occur until the concentration of water molecules is even either side of the membrane, meaning the water potential is the same on both sides.

Surface area to volume ratio

The larger the surface area to volume ratio, the more effective transport is. Single celled organisms have a large surface area compared to their volume and they can rely on diffusion alone to meet their needs. Larger multi-celled organisms cannot meet their nutrient need by diffusion alone, hence the need for transport systems and specialised surface areas where there is a large surface for diffusion to take place.

Assessment practice 5.10

1 Describe the structure of a biological membrane.
2 Explain the methods of transport across a partially permeable membrane.
3 Produce a table to compare and contrast the different methods of transport.

Further reading and resources

Annets, F., Foale, S., Hartley, J., Hocking, S., Hudson, L., Kelly, T., Llewellyn, R., Musa, I. and Sorenson, J. (2010) *Applied Science Level 3*. Pearson.

Boyle, M. and Senior, K. (2008) *Biology* (3rd edition). Collins.

Getting started

Thermodynamics is about the links between heat, energy and power. Although it started with engineers designing and improving steam engines, it now helps shape our understanding in almost every area of science – from the chemical reactions that provide the energy for life to the mystery of space and time and the origins and future of the universe.

What do you already know about heat and energy? Try writing down all the definitions and important equations you can think of. When you have completed this unit, come back and see what you can add to your summary.

C Thermal physics, materials and fluids

Thermal physics in domestic and industrial applications

This subject builds up in small steps. When you can put it all together, it is surprisingly powerful and answers important questions. So try to make sure you grasp each definition and law along the way.

Measurements

You will need to be able to recognise and use the SI (Système Internationale) units for the following important quantities. Each unit starts with a capital letter because it is the name of a scientist who did important work.

Energy: unit Joule (J)

Energy is closely related to work (see below) and they share the same unit.

Power, symbol *P*: unit Watt (W)

We commonly deal with large quantities of power, so you also need to be familiar with the following multiples:

- kilowatt (kW) = 1000 Watt
- megawatt (MW) = 10^6 Watt
- gigawatt (GW) = 10^9 Watt.

1 Watt = 1 Joule per second ($J\ s^{-1}$).

(Do not confuse with the kilowatt-hour (kWh), which is a unit of energy, not power. As there are 3600 seconds in an hour and 1000 watts in a kW, 1 kWh = 3,600,000 J.)

Temperature, symbol *T*: unit Kelvin (K)

In this subject it is often important to use the **absolute temperature** in equations. So you need to know how to convert from degrees Celsius (°C) – also known as °Centigrade – to Kelvin (K). (Note that there is no degree sign before Kelvin.)

When you see the symbol *T* in an equation you should always use the absolute temperature in Kelvin.

Key terms

Absolute temperature – or '**thermodynamic temperature**' is measured on a scale starting at **absolute zero.** (Symbol *T*, SI unit Kelvin (K)). Convert temperatures in °C to Kelvin by adding 273.15 K, which is the freezing point of water on the Kelvin scale.

Thermodynamic temperature – (symbol *T*) the modern way of defining absolute temperature, which does not rely on the existence of an ideal gas, but instead uses the amounts of heat transferred in idealised engine cycles.

Absolute zero – the lowest temperature an object can be cooled to, where all thermal energy has been removed from it, there is minimal particle movement and it is in its lowest possible energy state.

Pressure, symbol *p*: unit Pascal (Pa)

Notice that the letter '*p*' is used for a lot of things in thermodynamics; so it is important to be careful and consistent about how you write them. Capital *P* means power. Try to always use a lower case *p* for pressure. The unit symbol Pa has the extra 'a' to distinguish it from power and pressure.

1 Pascal = 1 Newton per square metre ($N\ m^{-2}$)

Work done

When objects interact, the forces between them can lead to energy being transferred. Sometimes the energy is stored in a useful form, e.g. in a spring or in a gravitational or electric field (this is called potential energy). Sometimes energy is due to the speed of a moving object (this is called kinetic energy). Alternatively, the thermal energy (heat) content of an object may be increased.

When you measure **work**, you are focussing on the amount of energy transferred during the process.

Key term

Work – the work done in a process is the amount of mechanical energy transferred. (Symbol: *W*, SI unit: Joule (J)).

(Be careful not to confuse *W*, for the energy transfer *quantity* 'work', with W meaning the power *unit* 'Watt'.)

Work done by a force

The most obvious way of doing work is to exert a force in order to move something, e.g. pushing a child on a swing, or lifting a bag of shopping onto a table. Calculating the work done in both those examples is straightforward:

> work done = force × perpendicular distance moved in the direction of the force ($W = F \times \Delta s$)

The symbol Δ means 'change in'. So the 'change in position', $\Delta s = s_2 - s_1$, where s_1 is the initial value of the position (displacement) and s_2 is its final value. Note that we always start with the final value and take away the initial value.

Forces occur in opposing pairs: so the force you exert to lift the shopping is opposed to the weight of the shopping itself. When you lift the shopping and it moves up in the same direction, you do positive work on it. But if you let it down again onto the floor, the work you do is negative – or to put it another way, the shopping then does work on your arm. Some old clocks operate in that way, with weights that slowly fall and drive the mechanism.

But sometimes the force exerted is in a different direction from the movement. For example, when you push something up a slope, the force you are working against is gravity. So, for calculating work done, it is only the vertical height gained that counts as the distance moved – the part of the movement that is along the line that the force acts.

Work done by a gas expanding

Another important way that work is done is when a gas expands to take up a larger volume. This is what happens in the cylinders of a steam engine or motor car engine (petrol or diesel). Hot, high pressure vapour or gas pushes against a piston and moves it. To calculate the force exerted by the gas you would need to multiply the gas pressure by the area of the piston head, *A*. So $F = p\,A$.

When the piston moves, the volume change for the gas is that same area, *A*, multiplied by the distance moved by the piston. So $\Delta V = A\,\Delta s$.

Putting those two facts together, you can now calculate the work done by the expanding gas:

> work done = pressure × volume change
> ($W = p \times \Delta V$)

Note that this definition of work saves you having to worry about directions. A gas exerts its pressure in all directions on every surface of the container that defines its volume.

Law of conservation of energy

Because forces always occur in pairs that are equal in size but opposite in direction (Newton's First Law of Motion), the energy *transferred to* an object when work is *done on it* is always equal in size to the amount of *energy lost* by the second object that is *doing the work*. So overall the total energy of the pair of objects remains unchanged. As this is true for every pair of objects in the universe, you can assume that the total energy of the whole universe must also be constant, whatever goes on within it.

This is known as the Law (or principle) of Conservation of Energy. It is one of the most fundamental laws of physics.

Thermodynamics is largely about simplifying things so that they can be studied and measured. You cannot study the whole universe all at once. So scientists define a **system** by drawing an imaginary boundary around what they are studying. Then they can talk about the system and its **surroundings**, i.e. the rest of the universe. (You will often study two systems and measure the interactions between them.)

Key terms

System – the part of the universe whose properties you are investigating. It is enclosed by a boundary defined by you, the experimenter.

Surroundings – the rest of the universe, outside the system boundary.

The energy contained in a system can be in many different forms, including those shown in Table 5.11.

Table 5.11 Types of energy and their nature

Type of energy	Nature of the energy
mechanical energy	either potential energy, due to its mass and position in a gravitational field, or kinetic energy, due to an object's mass and speed of motion
electrical energy	can be associated with static electric charge and potential, or with moving charges, current and magnetism
chemical energy	intrinsic to the microscopic structure of the material and chemical bond energies
nuclear energy	due to the binding of subatomic particles (protons and neutrons) in the nuclei of atoms – changes in that are what cause radioactive decay, and also where nuclear power comes from
thermal energy	due to the microscopic vibrations and movements of atoms and molecules in a material – movements that you cannot directly measure, but that give rise to a measurable quantity, temperature

Link

See *Unit 1: Principles and Applications of Science 1*, Learning Aim A, for more on bond energies.

The principle of energy conservation means that when, in some process, you observe a reduction in one kind of energy, you will always find an equivalent increase in other forms of energy. Energy is transferred, but never lost. (Nevertheless, sometimes when energy is transferred to the surroundings in a form that is not 'useful', people do talk about energy losses. You will think more about 'useful' and 'wasted' energy when calculating efficiency.)

Heat and temperature

When two systems interact, there may be a flow of **heat** between them due to a **temperature** difference. This is a transfer of thermal energy from one system to the other. Flow of heat between two bodies in thermal contact (i.e. they are not completely insulated from one another) will continue until they reach **thermal equilibrium** and have the same temperature.

Temperature is such a commonly experienced quantity that scientists did not realise they need to define it until after they had developed the first two laws of thermodynamics. So the definition of temperature is often called the Zeroth Law of Thermodynamics. That simply states that there exists a property called temperature such that any two bodies at the same temperature will be in equilibrium with one another, and indeed also in equilibrium with a third object – a **thermometer** – used to measure each of their temperatures – see Figure 5.22.

Key terms

Heat – the quantity of thermal energy transferred during a process.

Temperature – the physical quantity that determines the rate at which heat will flow – from a 'hot' body (system) to a 'colder' (i.e. lower temperature) one. It is directly proportional to the average kinetic energy of the molecules.

Thermal equilibrium – exists when two systems are in thermal contact, but there is no net transfer of heat because they are at the same temperature.

Thermometer – a device with a readily measurable property that varies directly with temperature.

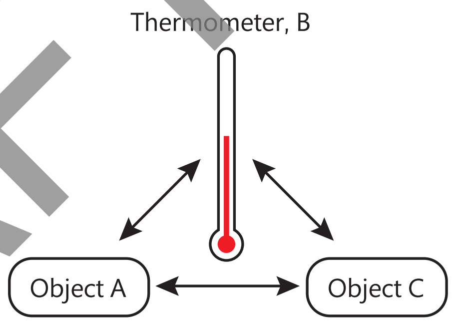


Figure 5.22 Thermometry – systems in thermal equilibrium

At a microscopic level, temperature is linked to the average kinetic energy of vibration and of motion of the atoms and molecules. You can picture that particles continually bump against one another and so exchange thermal energy, with the result that the energy gradually gets spread more and more evenly. At equilibrium, the rate of energy transfer in one direction will exactly match that transferred back in the other direction. So overall there will be no net transfer – no heat flow.

Calculating efficiency

Not all forms of energy are as useful and controllable as others. Mechanical energy is what you can most readily control, and it can be converted into electrical energy using a generator. Electrical energy can be converted into many other energy forms. But it is difficult to produce anything other than thermal energy from nuclear or chemical energy – fuel cells being one exception.

Every other energy form can eventually be completely converted into thermal energy. When you bounce a ball, stir a liquid, hammer a nail . . . the so-called 'lost' mechanical energy always turns up as an increase in temperature in those things. It is not actually lost. An electric kettle or a microwave oven is always 100% efficient at converting electrical energy into thermal energy, even though some of the heat from the process may be 'lost' in heating electrical cables or the air in the surroundings rather than in heating your food.

But the same is not true in reverse. You cannot take the warmth from your cup of tea or coffee and turn it back 100% into electricity or into mechanical energy. Thermal energy can only be made to do useful work if you have a temperature difference to create a heat flow; and at the end of the process a lot of the energy will still be thermal – just at a lower temperature.

As so much of the energy our society uses comes from heat – from the high temperature of a nuclear reactor or from burning a chemical fuel – improving as far as is possible the efficiency of engines that use heat is very important. For example, steam turbines used in power stations and the internal combustion engines of motor cars are **heat engines**.

The idea of **energy efficiency** is all about how much of the energy you use actually ends up in the useful form you are trying to create.

You can therefore define the **thermal efficiency** of a heat engine quite straightforwardly as:

$$\text{efficiency } (\eta) = \frac{\text{work output, } W}{\text{heat input, } Q_{\text{in}}}$$

This is because the purpose of the engine is to turn a source of thermal energy into useful mechanical work.

Key terms

Heat engines – machines designed to convert thermal energy into useful mechanical work.

Energy efficiency – is the fraction of the total energy input that is converted into a useful energy output. (The fraction may also be expressed as a percentage – just divide the top of the fraction by the bottom, then multiply by 100.)

Thermal efficiency – another name for energy efficiency when the energy input is in the form of heat. This name emphasises that the temperatures involved are of critical importance to the efficiency.

You can understand why this efficiency is always much less than 100% by looking at the diagrams of heat engines in Figure 5.23.

From these diagrams, you can see that heat engines always have a heat input, Q_{in}, at a high temperature, T_H, but they cannot avoid also having a heat output, Q_{out}, at a lower temperature, T_C. It is that temperature difference that makes the heat flow, and so drives the process in the engine – usually the expansion of a gas or vapour in a turbine or cylinder.

Using the principle of energy conservation:

$$\text{total energy input} = \text{total energy output}$$

So, in the case of a well-designed heat engine that produces useful work, with nothing wasted as friction or noise:

$$Q_{\text{in}} = W + Q_{\text{out}} \text{ and rearranging: } W = Q_{\text{in}} - Q_{\text{out}}$$

Therefore, the equation for the efficiency of the heat engine becomes:

$$\text{efficiency } (\eta) = \frac{W}{Q_{\text{in}}} = \frac{Q_{\text{in}} - Q_{\text{out}}}{Q_{\text{in}}} = 1 - \frac{Q_{\text{out}}}{Q_{\text{in}}}$$

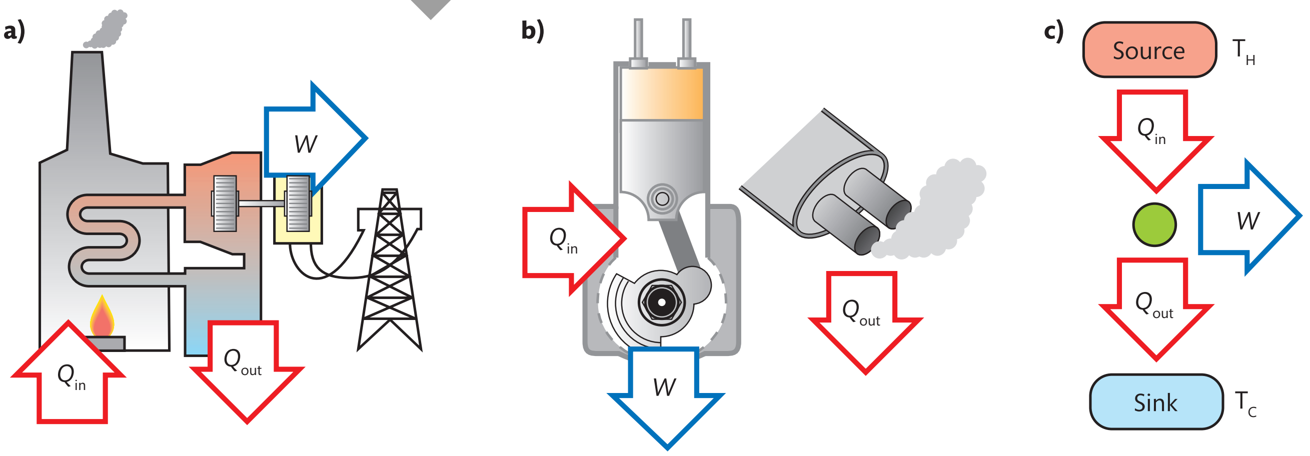


Figure 5.23 Energy transfers in a) a steam turbine b) a car engine c) a generalised heat engine

Nineteenth-century engineers, working to improve steam engines, began to realise that the efficiency they could achieve seemed to be limited by the temperatures involved, T_H and T_C. But to understand why, you will first need to develop your understanding of a few more concepts of thermal physics, which these engineers were the first to discover.

PAUSE POINT You have covered many of the key ideas and terms used in thermal physics. Close the book and test how many you can recall and explain.

Hint Try explaining how thermometry works, and what a heat engine does.

Extend Why is electricity such a useful form of energy? Why is it so expensive to produce?

Ideal gas equation

Solids, liquids and gases are all composed of atoms and molecules. In a gas, the molecules are separated from one another and are constantly in motion. This understanding of the nature of a gas is called kinetic theory ('kinetic' means 'moving') and it explains why the pressure, p, of a gas varies according to its temperature, T, and the number of molecules, N, that are contained in a volume, V.

If you make the approximations that, in an **ideal gas** (i.e. one that fits this simplified theory):

- the molecules are mostly so far apart that they do not exert any forces on one another, except when they actually collide
- their collisions are always elastic – i.e. they exchange kinetic energy, but none is lost overall
- the size of the molecules themselves is negligibly small compared with the volume the gas is occupying

then, from this theory, you can derive an **equation of state** for the gas:

$$pV = NkT$$

where k is the Boltzmann constant, which is defined statistically from the average kinetic energy of the gas molecules at a given temperature – average molecular kinetic energy $= \frac{1}{2}kT$ in each of the directions, x, y and z.

This ideal gas equation is useful, because at relatively low pressures and large volumes, and well above their condensation temperature (the temperature at which they become liquid), real gases show behaviour that is very close to the 'ideal'.

Key term

Ideal gas – a theoretical model of a gas, where the molecules are assumed to be point particles that take up no volume and that exert no forces on one another in between their elastic collisions. This simplified model makes it easier to calculate their behaviour, based on Newton's laws of motion.

Key term

Equation of state – an equation connecting the measurable quantities that define the physical state of a system; for example, the pressure, volume, temperature and amounts of material (numbers of molecules) present in the system.

It is also very helpful for developing the idea of a thermometer to measure temperature, because measurements of gas pressure and volume can give us a direct proportionality with absolute temperature, T. (See Figures 5.24 and 5.25.)

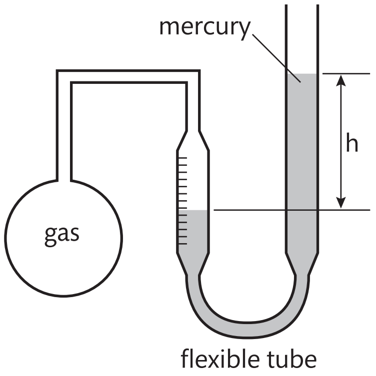


Figure 5.24 A constant volume gas thermometer; the gas expands on heating, but the mercury manometer is adjusted to bring it back to its original volume, and then a pressure reading is taken.

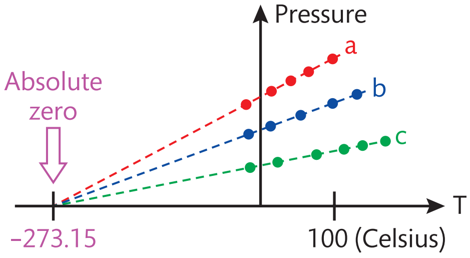


Figure 5.25 Extrapolating the readings taken from a constant volume gas thermometer indicates an intercept on the temperature axis at approximately –273 °C. This is how the existence of an absolute zero for temperature was first predicted.

Worked example: Work done by gas escaping from a balloon – the rocket principle

An inflated balloon has a volume of 5 litres and contains air at 120 kPa, which is higher than the surrounding atmospheric pressure of 100 kPa. If the air is allowed to escape rapidly, the balloon flies off like a rocket because the expanding gas is doing work. But there is no piston.

To estimate the work done by the gas, first calculate the volume change:

The temperature of the air at the end of the process will be the same as at the start so, using the ideal gas equation of state, $p_1V_1 = p_2V_2$. Thus the final volume of the gas after expansion is given by:

$$V_2 = V_1 \times p_1/p_2 = 5\,\text{litre} \times 120/100 = 6\,\text{litre}$$

So $\Delta V = \text{initial volume} - \text{final volume} = V_2 - V_1 = 6\,\text{litre} - 5\,\text{litre} = 1\,\text{litre} = 1 \times 10^{-3}\,\text{m}^3$

The escaping air pushes against the external atmosphere at 100 kPa.

So Work done, $W = p\,\Delta V = 100\,\text{kPa} \times 1 \times 10^{-3}\,\text{m}^3 = 100\,\text{J}$

The balloon itself has quite a small mass – just a few grams – so when 100 J is converted to kinetic energy using the equation ½ mv^2, it will equate to quite a high velocity.

Changes of state

When the temperature of a gas cools, and its pressure and volume reduce accordingly, the average spacing between the molecules becomes much less so that they exert significant attractive forces on one another. A point comes where those forces are sufficient to bind the molecules into a disorganised but quite closely packed structure – a liquid. This is **condensation**, and it happens at definite temperature and pressure values – the boiling point.

The molecules in a liquid vibrate and even flow round one another, but they do not have enough energy to break free of the bonding forces and to travel randomly over long distances as molecules do in a gas. The reverse process, when a liquid is heated and, progressively, more and more of its molecules gain enough kinetic energy to break free of the liquid surface, is called **vaporisation**.

Vaporisation requires a lot of energy input and, correspondingly, during condensation an equivalent amount of energy is given up by the condensing gas. This energy is called the **latent heat** of vaporisation – 'latent' because it appears to be used up without doing anything: there is no temperature rise during the process of vaporisation and no temperature fall during condensation. All the latent heat energy goes into the process of changing the state of the material from liquid to gas, or back again.

When a system contains more than one substance – e.g. when it also contains air – then the **saturated vapour** pressure in equilibrium with the liquid is only part of the total air pressure. Boiling occurs when the temperature increases to a point where the saturated vapour pressure (SVP) has risen to equal the atmospheric pressure. The boiling **vapour** then drives out the rest of the air, filling all the space. So rapidly boiling off a liquid in a confined space can easily cause asphyxiation and death.

A similar energy change occurs when temperature falls and pressure increases sufficiently for the material to form a more ordered and fixed structure – **freezing** to form a solid. The reverse process of solid melting to a liquid is called **fusion**.

Both the freezing point and the boiling point happen at different temperatures depending on the pressure of the system. Look at Figure 5.26 to see the regions of temperature and pressure in which the material exists as a mixture of two **phases** – solid and liquid or liquid and vapour. Because of these two-phase regions it is called a phase diagram.

Examining the phase diagram, you will see that there is one specific temperature and pressure at which three phases are in equilibrium – the solid state, liquid state and vapour state of the material: in the case of water these are ice, liquid water and water vapour (steam). This is called the **triple point** of water, and its temperature -273.16 K (0 °C) is the upper fixed point defining the Kelvin thermodynamic temperature scale, the lower fixed point being 0 K = absolute zero.

At the upper end of the liquid-vapour two-phase region is a maximum temperature and pressure called the **critical point**, above which there is just one phase, a relatively hot, dense supercritical gas.

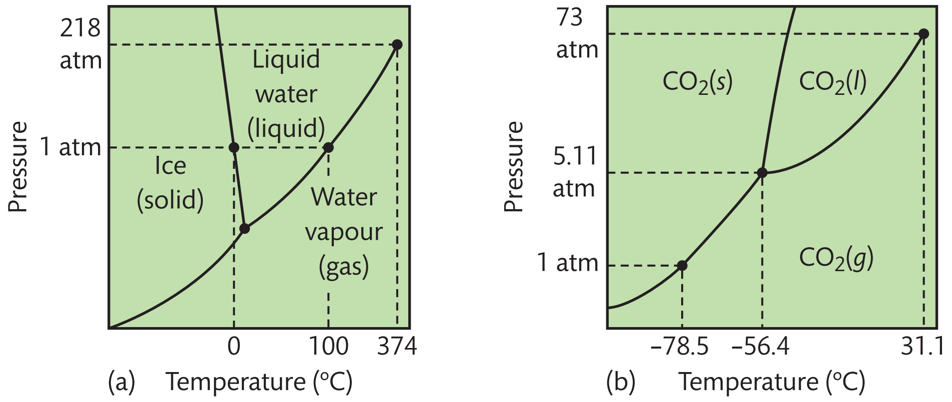


Figure 5.26 Temperature–pressure phase diagrams (a) for water and (b) for carbon dioxide.

Key terms

Latent heat – energy transferred that has the effect of changing the physical state of a substance without changing its temperature.

Vapour – is the name given to a gas that is at a sufficiently low temperature that it will change state to a liquid or solid if its pressure is increased sufficiently. In other words, a gas below its critical temperature (see below).

Vaporisation – is the change of physical state from liquid to gas. The reverse process is **condensation**.

Saturated vapour – is vapour that is in equilibrium with a liquid phase of the same substance. Any further compression causes more liquid to condense. The vapour cannot exert a higher pressure at that temperature than its saturated vapour pressure (SVP).

Fusion – (or melting) is the change of physical state from solid to liquid. The reverse process is **freezing**.

Phase – a separate part of system which is not uniform throughout. For example, a mixture of ice and water is a two-phase system, but a solution of salt in water is a one-phase system.

Triple point – the temperature and pressure at which all three phases: solid, liquid and vapour, can exist in equilibrium.

Critical point – the temperature and pressure at which the density of the liquid and vapour phases become identical. Above that critical temperature, only a supercritical gas phase can exist.

Theory into practice

Measuring temperature: Using a container of gas as a thermometer and measuring its changes of volume or of pressure is cumbersome. Moreover, there is no such thing as a perfectly ideal gas. As gases and vapours cool towards their condensation points, the distances between molecules become smaller and the forces exerted between them increase. So they begin to deviate from the Ideal Gas Equation.

So, in practice we define practical temperature scales that use a series of fixed temperature points defined by physical properties of a material – e.g. the triple point of water – and some good practical thermometers to divide the scale in between those points. Look up the International Temperature Scale of 1990 (IPS-90) to see the fixed points and thermometers specified for really accurate temperature measurement.

Internal energy and the first law of thermodynamics

If you take an object, like a tennis ball, you can lift it or throw it or hit it with a racquet to give it more mechanical energy – i.e. potential energy and kinetic energy. But that is only a small fraction of the total energy that the ball has. Inside it, and invisible to you, all the atoms and molecules are in thermal motion. Moreover, those particles are bound together by chemical bonds, and the nuclei of the atoms contain vast quantities of nuclear binding energy.

All the hidden energy inside a system that you are studying is called its internal energy, denoted by the letter *U*.

Of course, it is impossible to measure the total value of *U* in any direct way. What you can measure is the amount of energy transferred into or out of the system during any process – transfers of heat, *Q*, or work, *W*. From those definite measured quantities, you can calculate how much the internal energy must have changed:

$$\Delta U = Q - W$$

This equation is known as the first law of thermodynamics. It is just a statement of the principle of conservation of energy as it applies to thermal systems – systems that use and transfer heat.

In the equation:

- Q has a *positive* sign because we count heat input to the system as positive and heat transferred out as negative
- W has a *negative* sign because we count work as positive when the system does work for us and so gives up some of its internal energy.

The symbol Δ means 'change in', i.e. (final value) – (initial value). So ΔU means 'change in internal energy'.

Internal energy, U, is a function of the state of a system. This means that if the pressure, temperature and volume of each component in a system is known, then there must be a fixed amount of internal energy, U_1, in the system, associated with those values. And so, when there is a change in those variables that define the state of the system, there will be a corresponding change, ΔU, to a new fixed amount of internal energy, U_2. You can calculate the change, ΔU, from the measured values of Q and W, but you never know the absolute values of either U_1 or U_2.

(Note that Q and W are measured quantities of energy transferred, not changes in anything. So, correctly, you should not use the symbol Δ with them. However, doing so is a common mistake that you will even see in some books. So if you see that, just ignore the Δ.)

Natural processes and the second law of thermodynamics

Some things naturally tend to happen, and others never do. For example, a hot cup of tea or coffee gradually cools down, losing heat and slightly warming its surroundings. But you never see a warm drink absorbing heat from its cooler surroundings and heating up.

Similarly, if you put sugar into your hot drink it will gradually dissolve and mix naturally, though you might stir it to hurry the process along. But, once mixed, you never see all the sugar separating out again as crystals. It takes a lot of effort and energy input to concentrate a sugar solution and to get it to crystallise, and even then some of the sugar remains in solution.

If you watch a film or video running backwards, some things appear normal – they are reversible in time; but many other things are obviously wrong. A collapsed wall will not rebuild itself naturally. A spilt drink will not collect itself up and flow back into the cup. You can easily spot that time has been reversed in the film played backwards.

These everyday observations are the basis of the second law of thermodynamics, which can be stated in several different but equivalent forms.

- A natural process can never be reversed in its entirety.
- It is impossible to completely change heat into work.
- Heat will not flow from a colder body to a hotter one without an input of work.
- You cannot reverse the direction of time.

The second and third of the statements above apply the Second Law to heat engines and to heat pumps respectively, and you will be learning more about them in a later section.

But everything that happens around you, and indeed in your own body, is a result of **natural processes**. Nothing is ever fully reversible. Nevertheless, all the other laws and equations of physics – including the First Law of Thermodynamics (energy conservation) – are written as if time could be reversed. So it is only by using the Second Law alongside these other laws that you can determine which way things will go. This means the Second Law is particularly important for:

- engineers designing engines and machinery
- chemists working out which reactions will happen, and where the position of equilibrium will lie
- biologists understanding what drives life processes
- cosmologists exploring the past and predicting the future of the universe.

Key terms

Natural process – any real process that is not driven by an input of work or heat from a source outside the system being studied, but just occurs naturally.

Entropy – 'potential for change'. One way to calculate the entropy of a system would be to count how much randomness it contains – i.e. the number of equivalent energy states and positions its particles could be arranged in. The more ordered and compact and high in temperature something is the lower is its entropy. Things naturally become more randomly mixed up, more spread out and cooler – i.e. their entropy increases. So, at its start, the universe must have had a very low entropy indeed.

Link

Entropy can also be defined by using the concept of ideal heat engine cycles.

PAUSE POINT Close your book and write down definitions of the First and Second Laws of Thermodynamics. Think about how those two laws work together to explain what happens in the world around you. What drives the direction of industrial processes and the biochemical reactions in living organisms?

Hint One system (e.g. our sun) increasing in entropy may release energy to drive another system (e.g. plant leaves) which then creates some more ordered, lower entropy structures. But what happens overall to the order and entropy of the universe?

Extend Highly unusual things – coincidences – happen all the time. But can *you* repeat them? What does the Second Law mean when it states that certain things are 'impossible'?

Reversible processes (isothermal and adiabatic)

Having just said that no process is ever fully reversible in time, the *ideal* of a fully **reversible process** is a very useful idea.

There are two sets of conditions which, if fully achieved, would make a process reversible:

Heat transfer at constant temperature – an isothermal process

When two bodies are at thermal equilibrium, heat could move in very small amounts either way across a boundary between them. So that would be a reversible process.

But in reality, to make the heat flow there always has to be a small temperature difference. So an **isothermal** process is an ideal that you can get close to but never quite achieve. Nevertheless, you can use the *idea* of a reversible isothermal process to think about the limits to achieving greater efficiency.

Slow heat transfer processes with excellent thermal contact, achieved in a large heat exchanger, are close to being isothermal.

No heat transfer at all – an adiabatic process

If no heat is transferred, and work is done with no conversion of energy into heat, then all the work will result in increased potential or kinetic energy. That energy could then be used to reverse in its entirety the work previously done – a fully reversible process, which could go on being repeated for ever and so create perpetual motion.

But it is never possible to thermally isolate a system completely. If its temperature rises above, or falls below, the temperature of its surroundings, it will transfer some heat. Even the best insulation gives some heat leakage. Also, any practical machine for doing work always suffers from some kind of frictional or viscous energy conversion into heat. Once again, it is the *idea* of a reversible **adiabatic** process that you will find useful – an ideal to strive for.

Very quick compression or expansion in a machine with low friction losses is close to being adiabatic.

Key terms

Reversible process – one which could be fully reversed in time, following the same path and exactly reversing the quantities of heat transferred and of work done. This is an ideal concept that can never be fully achieved in practice.

Isothermal – happens at one fixed temperature. To be reversible, an isothermal process would have to have a negligibly small temperature gradient across the system boundaries, so would be extremely slow.

Adiabatic – means 'no transfer'. It describes a process in a system that is not just well-insulated but is totally thermally isolated, so there is zero heat transfer. Also all work done in the process must be friction free and not create any extra thermal energy.

Idealised engine cycles

Now that you have the idea of two kinds of reversible process, you can go on to imagine a perfectly reversible engine for converting heat from a high temperature heat source into work. This *ideal* engine would be more efficient than any real engine you could build. It tells you the limits of what could be possible.

The ideal cycle is named after its inventor, Carnot, and comprises four processes for a working fluid – e.g. a gas or a liquid-vapour mixture:

1–2 adiabatic compression – zero heat transfer.

2–3 isothermal expansion – heat absorbed by the system, Q_{in}

3–4 adiabatic expansion – zero heat transfer

4–1 isothermal compression – heat absorbed by the system, $-Q_{out}$

Completing the four processes returns the working fluid to its original state, but work has been done, which can be measured from the areas under the pressure-volume curve ($W = \Sigma\, p\, \delta V$ where δV signifies a minute change in volume):

- over the two expansion processes: work done *by* the system counts positive, W_{out}
- over the two compression processes: work done *on* the system counts negative, $-W_{in}$.

So the net work done in the cycle is $W = W_{out} - W_{in}$. That is equivalent to the area on the *p*-*V* diagram (see Figure 5.27) enclosed by the cycle of processes.

Maximum thermal efficiency – the efficiency of an ideal Carnot cycle engine

You have already seen that the efficiency of a heat engine producing useful work from a heat input is given by:

$$\eta = \frac{W}{Q_{in}} = \frac{Q_{in} - Q_{out}}{Q_{in}} = 1 - \frac{Q_{out}}{Q_{in}}$$

Now you can apply this general formula to calculate the efficiency of the idealised 'Carnot' engine cycle.

What is the ratio of the heat input to the heat output, (Q_{out}/Q_{in})? How is it affected by the temperatures, T_H and T_C, at which those two heat transfers occur?

Well, imagine that the working fluid for your Carnot cycle were an ideal gas, obeying the equation: $pV = NkT$.

During the two isothermal processes:

- $\Delta U = 0$, i.e. the amount of thermal energy in the gas stays constant because the temperature is constant. So, applying the First Law equation, $\Delta U = Q - W$, you can see that: the heat transferred, $Q = W$, the work done, which equals the area under the *p*-*V* curve
- *pV* is constant for an ideal gas. So the process follows a *p*-*V* curve where *p* is proportional to $1/V$.

During the two adiabatic processes:

- $Q = 0$. So, again applying the First Law equation, you get: $\Delta U = -W$, which again equals the area under the *p*-*V* curve.
- For an ideal gas, **heat capacity** at constant volume, $C_V = \Delta U / \Delta T$, is a constant value because all the energy supplied goes into raising the average kinetic energy of the gas molecules, and thereby also raising the gas temperature. So that equation, combined with the ideal gas equation, fixes the shape of the *p*-*V* curve and makes it possible to calculate the volume changes in the adiabatic processes.

Key term

Heat capacity – the number of heat units needed to raise the temperature of a body by one degree. Also known as thermal capacity.

Doing the maths to sum up the areas under those curves for an ideal gas gives a simple relationship between the volumes at the four points of intersection between the four process curves. That, in turn, leads to a surprisingly straightforward relationship between the temperature of each isothermal in a reversible cycle and the amount of heat transferred reversibly in that process:

$$\left(\frac{Q_{in}}{Q_{out}}\right)_{rev} = \frac{T_H}{T_C}$$

This important result does three things:

1. It means you can rewrite the formula for **maximum theoretical efficiency** of a heat engine using absolute temperatures instead of amounts of heat:

$$\eta_{rev} = \frac{T_H - T_C}{T_C} = 1 - \frac{T_H}{T_C}$$

2. It gives a new way to define absolute temperature – by saying that thermodynamic temperature, *T*, varies in proportion to the amounts of heat transferred in an ideal engine cycle. That is a definition that does not depend on the existence of an ideal gas. The Carnot ideal cycle could equally well be performed with a real gas or with any other working fluid. In fact, the cycle of isothermal and adiabatic processes can also be performed with electrical or magnetic systems, rather than using a fluid of any kind.
3. It gives the idea that the ratio $(Q/T)_{rev}$ has a fundamental meaning – that heat transferred at a higher temperature is intrinsically more useful, and that this usefulness can be measured using a quantity derived from that ratio, which is called entropy.

Key term

Maximum theoretical efficiency (symbol η_{rev}) – no real engine can achieve a thermal efficiency higher than that of an ideal reversible engine operating between the same temperatures. (If this were not so it would be possible to create a machine that violated the Second Law and moved heat from a colder to a hotter body without doing work.) In practice, engine efficiency is always lower than this because all real engines have irreversible processes that produce less net work output and instead output more heat at the low temperature.

PAUSE POINT 'Laws' apply generally and can be used in real situations. 'Ideals' are simplified ideas and equations that only almost work under specified conditions. Use examples from thermodynamics to help explain this difference between laws and ideals.

Hint Examples of ideals include: Gas equation of state; Adiabatic changes; Isothermal changes. Explain why each of these is never exactly true.

Extend Also explain why each of those 'ideals' is nevertheless a useful idea.

Heat engines, refrigerators and heat pumps

Real heat engine cycles

Engines are an important part of modern life and their efficiency is a matter of concern, firstly because fuels in general are a limited resource and secondly because the burning of fossil fuels releases carbon dioxide, CO_2, and hence contributes to global warming and climate change.

Steam turbine engines

The majority of electrical power worldwide is produced by generators that are turned by a steam turbine. Water is heated in a 'boiler' heat exchanger to produce steam, and the steam is then expanded through a turbine, with the low pressure steam being cooled and condensed back into water. Cooling the steam quickly in a heat exchanger reduces its pressure to below atmospheric, and the partial vacuum drawn by this cooling increases the pressure difference that drives the turbine. Because liquid water is nearly incompressible, only a small amount of energy is needed to pump it back up to the pressure of the boiler inlet.

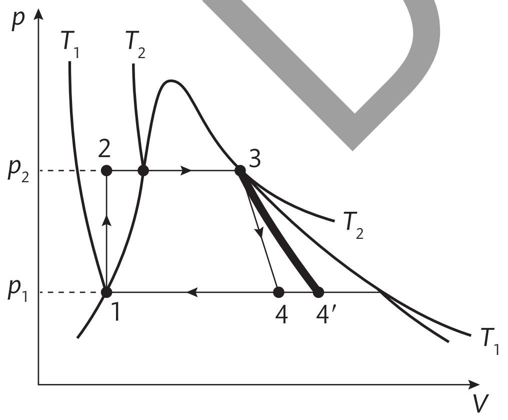


Figure 5.27 The Rankine cycle of a steam turbine engine shown on a water/steam *p-V* diagram

This practical engine cycle is named after its Scottish inventor, Rankine. The water (working fluid) operates in a closed cycle, as shown in Figure 5.27, with condensed steam being returned as liquid water to the boiler. But the heat rejected from the condensing steam and transferred in the condenser heat exchanger to a supply of cooling river or sea water, makes that cooling water hot. It is cooling that down again so that it is ready to be returned to the river or sea which causes the clouds of steam and water droplets that are seen escaping from massive cooling towers.

The high specific latent heat of vaporisation of water, $2.26\,\text{MJ kg}^{-1}$, has the advantage that large quantities of heat energy are transferred by moving only a relatively small mass of water, keeping machinery sizes compact. But it also limits the temperature at which most of the heat is input, which in turn limits the thermal efficiency obtainable – typically up to only about 42% efficiency. Even this figure is only achieved when the steam continues to be heated after it has boiled, which is called 'superheating'. This makes the superheated steam 'dry' (i.e. there is no trace of liquid water) and at a higher temperature than the boiling point.

Nuclear power stations use saturated steam with no superheating (straight from the boiler and 'wet' because it is in contact with liquid water). So they have even lower thermal efficiencies of around only 30%. But efficiency is less of a concern for them than safety of containment of radioactive hazards. They are not emitting CO_2 into the atmosphere to add to global warming, and the supply of heat from their fuel is plentiful.

To keep the heat exchange processes efficient – i.e. near reversible – the size of the heat exchangers used with steam engines needs to be very large indeed, so steam turbine systems are only practical for stationary engines or for powering large ships; they are much too large and heavy for use in road vehicles.

Internal combustion engines – petrol or gas, diesel and gas turbine

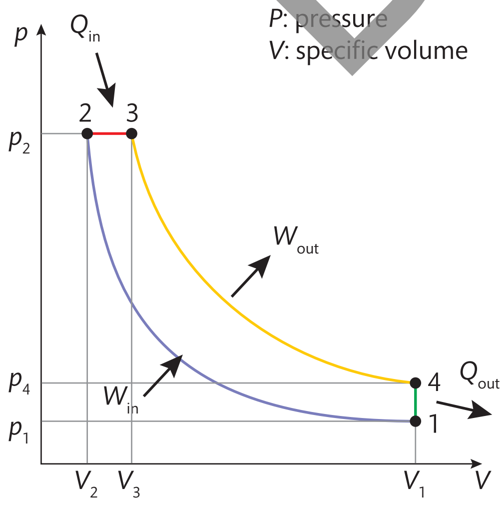
The working fluid in these engines is mostly air. Air is sucked or pumped in and then compressed, with some fuel added. Combustion inside the engine heats the gas very rapidly indeed, and then that mixture of hot air and combustion products does work as it expands. Finally, the gases are exhausted out to the atmosphere (see Figure 5.28).

Piston engines, used in motor vehicles, are small and lightweight, but they are mechanically complicated, creating friction losses, and it is difficult to achieve fully effective combustion in such a short space of time. On the other hand, the high temperatures at which the heat is supplied give them the potential for higher thermal efficiencies. The reason diesel engines give the higher efficiencies is because the air can be more highly compressed, giving a higher initial temperature, at which point fuel is injected and ignites immediately. Petrol is much more flammable, so compression ratios in petrol engines have to be kept fairly low to avoid problems with 'pre-ignition'.

Higher thermal efficiency makes diesel engines burn less fuel and produce less CO_2 than petrol engines. But incomplete combustion in diesel engines creates other problems of air pollution in cities.

Liquefied natural gas is a tricky fuel to handle in vehicles because of the high pressure containment needed, but it can combine clean complete combustion with fairly high thermal efficiency.

Gas turbine engines also use internal combustion (i.e. burning the fuel inside the engine), but the ignition and the hot gas expansion happen in a turbine, instead of a piston engine. Turbines are mechanically efficient, directly producing the rotational motion needed for electricity generation, and, with exhaust gas recycling included, they can become the most efficient and clean burning engines of all. But to achieve this, they must have large heat exchangers and they need to be running continuously at close to full power. So they are widely used in power stations that provide the main electricity supply load – i.e. the base level of power demand that is used all day and all night long. Other types of power station, which can be switched off and on easily or run on reduced load, are needed to provide for the surges in electricity demand at peak load times like early mornings and evenings, when everyone wants to turn lights on, cook and heat their homes.



▶ **Figure 5.28** The Diesel cycle, shown on an air pressure-volume diagram

Muscle power

The chemical processes operating in biological systems are complex, but it is still possible to investigate the thermal efficiency of muscle tissues, measuring their energy conversion into useful work. Results show that, once again, their efficiency depends strongly on the temperature at which a muscle is working. Of course, the temperature differences are very small compared with the numerical size of the absolute temperature, so the maximum theoretical efficiency of an ideal Carnot engine under those conditions would be very low, and the actual practical efficiency values are even smaller.

Being warm-blooded makes a huge difference to the power that an animal's muscles can deliver. Equally, taking away the heat that is generated by 'burning' foods like sugars and fats inside body cells is also essential to keeping a body working efficiently. That's why humans perspire and dogs pant. Precisely because living organisms are so diverse and complex, simplifying a subject of interest and looking at it as a 'system', with inputs and outputs of energy and materials, often a powerful approach in biology.

Reverse cycles

You have been thinking a lot about idealised reversible cycles. So, what happens if you do reverse an engine cycle and run the machinery and the heat transfer processes in the opposite direction?

The reverse carnot cycle

1–2 Adiabatic compression raises the working fluid to a higher pressure and the temperature, T_H.
2–3 Isothermal compression then progressively rejects heat, Q_{out}, to the surroundings at that high temperature.
3–4 Adiabatic expansion next allows the still hot but compressed fluid to reduce pressure and cool rapidly to a new lower temperature, T_C.
4–1 Isothermal expansion finally has the fluid absorbing heat, Q_{in}, at the low temperature and gradually expanding until it is in its original condition and ready to start the cycle again.

This is a machine for moving heat from a colder body to a hotter body, as shown in Figure 5.29, but it requires an input of work, W_{in}, to perform the adiabatic compression, which is larger than the amount of work, W_{out}, that could be recovered in the adiabatic expansion process. The net work done in the cycle is $W = W_{in} - W_{out}$.

Everything is back to front compared with the equations for the normal engine cycle. The equation for W is reversed because you are now focussing on the work done *on* the system, i.e. an input of work that would have a negative sign.

ΔU, for the whole cycle must be zero, so using the First Law equation: $\Delta U = Q - W$, simply gives $Q = W$.

This means that Q for the cycle must also be negative – i.e. there is a net quantity of heat given out. In other words Q_{out} is bigger than Q_{in}.

Q_{in} is now associated with the low temperature, T_C, and Q_{out} with the high temperature, T_H. There is more heat given out at the high temperature than is absorbed at the low temperature.

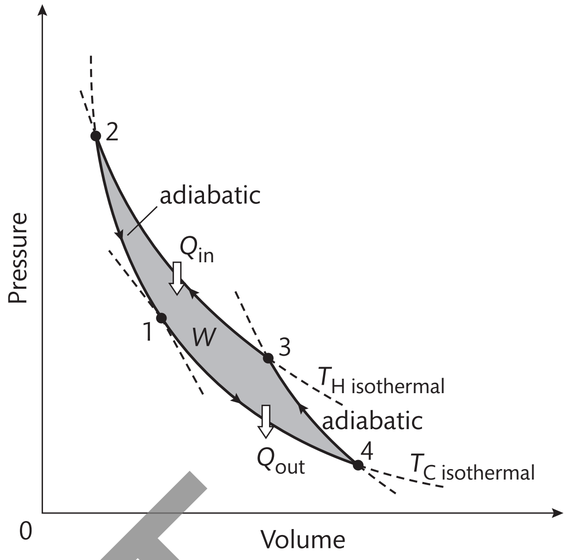


Figure 5.29 A reverse Carnot cycle removes heat from a low temperature reservoir, has an overall input of work equal to the area enclosed by the cycle, and rejects all that energy as heat into the higher temperature reservoir.

Vapour compression refrigerators and heat pumps

The main components of a refrigerator or heat pump are shown in Figure 5.30. For a real cycle to move heat from cold to hot, you would usually choose a liquid and condensable vapour system as the working fluid, rather than just a gas. That is because latent heat of vaporisation is helpful to the cost-effectiveness of these reverse cycle machines in several ways.

- Condensation and evaporation in the two heat exchangers are naturally isothermal processes, and hence closer to reversible, because the latent heat stops the temperature rising or falling during those changes of state.
- That in turn keeps the temperature difference between the two heat exchange processes as small as possible. While a small temperature difference limits the efficiency of an engine, it maximises the coefficient of performance (CoP) of a refrigerator or heat pump – see below.
- A high specific latent heat of vaporisation minimises the quantity of fluid that has to be moved round the system, and so keeps the size of the compressor equipment smaller and its price lower.
- Liquids are nearly incompressible – i.e. they have virtually no volume change when the pressure on them

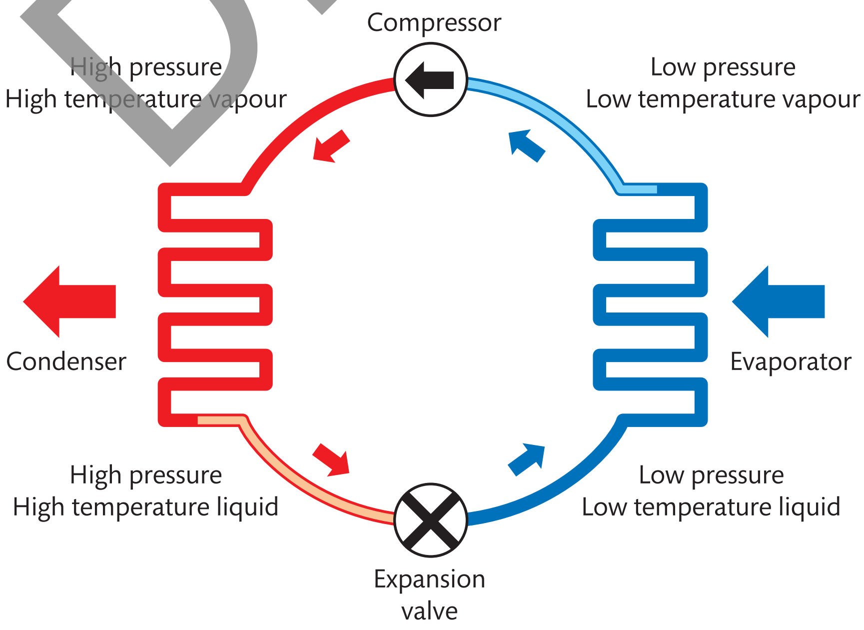


Figure 5.30 The main components of a refrigerator or a heat pump

is dropped. So the work done, $W = p\,\Delta V$, is very small indeed. This means it is not necessary to manufacture and use a turbine to recover that work. Just allowing liquid to drop at pressure through a simple valve, or in small domestic systems, using only a length of capillary tube, comes close to achieving the adiabatic expansion part of the cycle with only a very small amount of potential work output being wasted. At the new lower pressure, the liquid immediately begins to boil and in doing so begins to absorb latent heat. But by that time it is already in the low temperature heat exchanger – known as the evaporator – undertaking the isothermal expansion part of the cycle.

PAUSE POINT Check your understanding of the systems - engines, heat pumps and refrigerators, and even muscles – that you have just studied. They all operate a cycle of processes. Explain why that is the best way to think about them.

Hint Think about the working fluid in each, where it comes from and where it goes to.

Extend Heat pumps produce more heat energy output than the electrical energy input they use. Explain why that doesn't violate either the First Law or the Second Law of Thermodynamics.

Maximum theoretical coefficient of performance (CoP)

In the response to global warming and climate change, heat pumps and refrigerators are becoming increasingly important. Heat pumps can help us save energy on heating bills. But, on the other hand, refrigerators, freezers and air conditioning need a lot of electric power to keep them running. You therefore need to know how to rate these systems' performance and hence choose the most effective solution.

Theory into practice

Demand for cooling: Changes in our climate coupled with economic growth in developing countries is rapidly increasing the need for efficient cooling technologies. It has been predicted that global air-conditioning demand will grow 33-fold during this century, and that by 2060 it will consume more energy globally than is used for heating. Refrigeration is essential for many aspects of civilisation: food, medicine, energy, data and industry.

But cold is highly polluting. Already, around 10% of global CO_2 emissions are caused by refrigeration and air-conditioning, and escapes of traditional refrigerant gases also contribute strongly to global warming. We need to find new ways to reduce demand for cold, to create it more efficiently and to use more off-peak and 'waste' energy, or indeed 'waste' cold.

For more ideas on what we can do to build a more sustainable future for the world, read *Doing Cold Smarter*, Gavin Harper and David Strahan, Birmingham Energy Institute, 2016.

The term 'coefficient of performance' is used, rather than 'efficiency' because the value is generally greater than 1. These machines move more thermal energy than the amount of energy (usually electrical) that they consume doing the work of compression. That is why heat pumps are so popular – they can be big energy savers.

$$\text{Coefficient of performance} = \frac{\text{Heat usefully moved}}{\text{Work input required}}$$

CoP for a heat pump

In a heat pump, the useful heat is the amount, Q_{out}, rejected at the high temperature, T_H. So:

$$\text{CoP (heat pump)} = \frac{Q_{\text{out}}}{W} = \frac{Q_{\text{out}}}{Q_{\text{out}} - Q_{\text{in}}}$$

And so the maximum theoretical CoP for an ideal Carnot cycle heat pump would be:

$$\text{CoP (heat pump)}_{\text{max}} = \frac{T_H}{T_H - T_C}$$

This is the exact inverse of the maximum theoretical efficiency of a heat engine.

Exam tip

So, in an exam you can find the maximum theoretical efficiency equation on the formula sheet, turn the formula upside down and use it to answer heat pump CoP questions.

If the temperature difference ($T_H - T_C$) is kept small, the CoP can become quite a large number, meaning a large multiplication factor for the heat delivered compared with the electrical energy consumed – so it can be a big energy saver.

But real heat pumps are not fully reversible; their compressors have frictional and viscous losses, their heat exchangers are not really isothermal but have temperature differentials to drive the heat flow, and the work done by the expanding liquid as it passes through the valve or capillary tube is not recovered. So the CoP of a real heat pump will always be lower than the theoretical maximum.

CoP for a refrigerator

In a heat pump, the useful heat is the amount, Q_{in}, absorbed at the high temperature, T_C. So:

$$\text{CoP (refrigerator)} = \frac{Q_{\text{in}}}{W} = \frac{Q_{\text{in}}}{Q_{\text{out}} - Q_{\text{in}}}$$

And so the maximum theoretical CoP for an ideal Carnot cycle heat pump would be:

$$\text{CoP (refrigerator)}_{\text{max}} = \frac{T_C}{T_H - T_C}$$

By comparing the two equations for CoP, you can see that:

$$\text{CoP(refrigerator)} = \text{CoP(heat pump)} - 1$$

Exam tip

Again, this makes it easy to answer CoP questions. For refrigerators just work out the heat pump CoP and then subtract 1.

So, for example, a chilled food cabinet refrigeration system in a super market, operating with a CoP value of 2.5, if also adapted to deliver warm air into the shopping space, could be viewed as a heat pump with a CoP value of 3.5.

Maximising CoP

Of the various departures from an ideal cycle in a real system, the easiest to tackle is the size of the heat differentials in the heat exchangers. Simply by using an oversized condenser and an oversized evaporator, the temperature differences can be kept low between the refrigerant (i.e. the working fluid) and the heat exchange fluid (usually either air or water) in those heat exchangers – to just one or two kelvin for example. Another way of keeping temperature differentials low is better circulation of the heat exchange fluid. But care has to be taken not to expend too much energy driving the circulation of the air or water, as that adds to the cost of running the plant and so will bring the overall effective CoP down.

Keeping to a minimum the temperature difference that the heat pump or refrigerator has to work across is the key to increasing the maximum theoretical CoP. So, if you are installing a heat pump system, you should look for a low temperature heat source that will give you the highest possible working temperature through the periods when your heat pump will do most of its work. Similarly, a refrigeration system needs a reliable coolant for its condenser during the hottest times of the year.

For that reason, simple air-to-refrigerant heat exchangers are often not the best option. Air temperature rises in summer, when refrigerators have to do most of their work. And in winter, when heat pumps need to be most effective, the outside air temperature can fall very low. Water therefore often makes a better heat exchange fluid. Water towers are often used to provide cool water to the condenser of an air-conditioning or large refrigeration plant, even during hot summer days. Some of the water is allowed to evaporate as it falls through an air stream in the cooling tower, and the loss of latent heat of evaporation cools it before it is fed into the water-cooled condenser of the refrigeration system.

Underground coils carrying water can provide a steady heat reservoir at around 8 °C all the year round. This is good for a heat pump as a heat source during the winter months, when outside air can be much colder than that. But it can also be good for cooling the condenser of an air-conditioning system during the summer. With appropriate design and the use of valves, an air-conditioning system can be reversed to become a heat pump, providing cost effective winter heating to the same building.

If a refrigerator has to use an air-cooled condenser, e.g. for a domestic fridge or freezer, then the siting of the equipment is important. Too often they are placed very close to a wall and under a work-top, where there is limited natural air circulation. The ambient air temperature around the condenser can then become much higher than desirable for efficient operation. The consequent raised temperature and pressure of refrigerant inside the condenser can lead to cut-out or complete failure of the equipment on a hot summer day.

Assessment practice 5.11

Figure 5.31 shows the results of an experiment in which a 1.5 kW kettle was filled with 1.2 kg of water, placed on a digital balance, and a temperature probe monitored the water temperature from the point at which the kettle was switched on.

a) Explain why there is a change in both graphs after 5.5 minutes.

b) Calculate from the data on the graphs a value for the specific latent heat of vaporisation of water

c) In a second experiment, where only 0.6 kg of water was put into the same kettle at the same starting temperature, the change point in the two graphs occurred after only 3.2 min. Use these two results to calculate a value for the specific heat capacity of water.

d) Suggest why both these results may be a little lower than the accepted literature values.

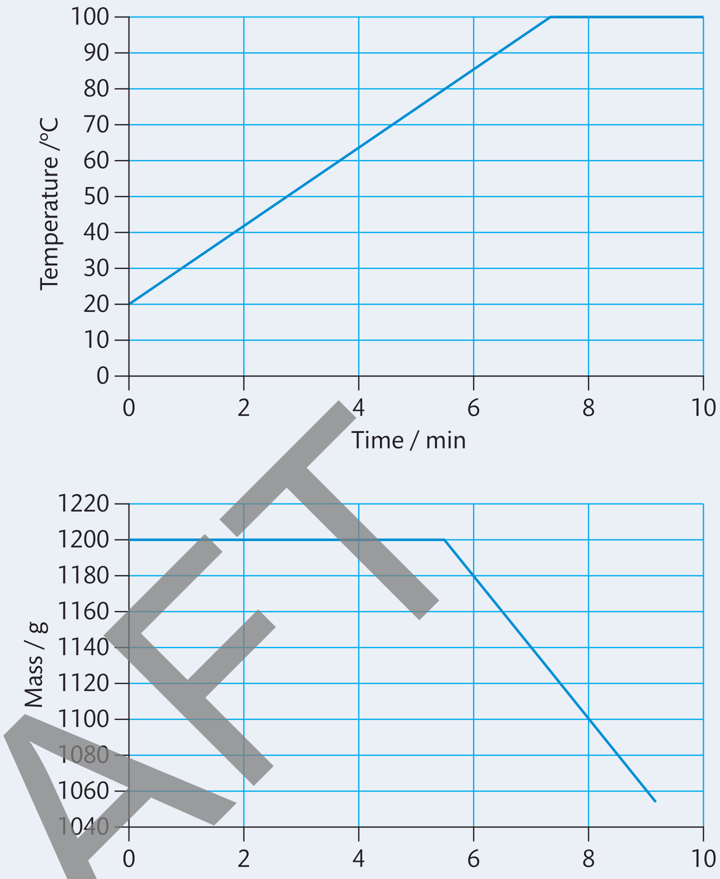


Figure 5.31 Heating water in a jug kettle

Assessment practice 5.12

A power station steam turbine which drives a 150 MW generator is fed by a boiler that burns fuel to give a heat input of 998 kJ every hour producing superheated steam at 565°C. The low pressure steam output from the turbine is condensed at 30°C. The generator is 99% efficient at converting mechanical work into electrical energy.

a) Calculate the maximum theoretical thermal efficiency of an ideal engine working between these temperatures.

b) Calculate the actual thermal efficiency of the steam turbine and explain what factors make it less than ideal.

c) Calculate the quantity of heat that the condenser unit must remove each hour and the mass of condensate that will be produced in the same period. (The latent heat of vaporisation of water is 2.26 MJ kg^{-1}.)

Assessment practice 5.13

A supermarket operates chilled food cabinets with an evaporator temperature of 2°C and a water-cooled condenser at 15°C.

a) Calculate the maximum theoretical coefficient of performance (CoP) for a refrigerator operating between these temperatures.

b) A suggestion has been made that the condenser heat output could be used to provide heating to the shopping space, using a heat exchanger operating at 30°C. Calculate the impact of doing this on the refrigeration CoP.

c) Explain how, alternatively, a two-stage system, with a heat pump taking its heat input from the condenser of the refrigeration system, might offer more control and flexibility across the seasons.

Materials in domestic and industrial applications

'Materials Science' is about understanding the properties of the *solid* materials you use to construct things, so that you can make them cost-effective, safe and reliable.

Elasticity and Hooke's Law

Solids have a shape and size that do not change unless a sufficient force is applied. Elastic behaviour is when a solid material is able to regain its original dimensions after the applied force is removed. If a shape or size change becomes permanent it is called a **plastic deformation**.

Solid materials are generally elastic – i.e. they can bounce back to shape – up to a certain limit, known as their **elastic limit**. After that point the material will either fail (break) or it will undergo plastic shape change.

Tensile experiments – i.e. testing materials by stretching them – were carried out originally by Sir Robert Hooke, and have been performed regularly since then by materials scientists. These experiments show that, for metals and many other solids over much of their elastic region, the relationship between applied force, F, and extension (increase in length), Δx, is a linear one. That is:

$$F = k\,\Delta x$$

The constant of proportionality, k, depends not only on the material used, but also on the original dimensions of the sample tested. So, while you may choose to use the equation in that form for a specific object – e.g. a spring of a certain size and specification, and to measure k, the object's spring constant; in general it is much more useful to write the law in terms of tensile stress and strain:

$$\tau = E\,\sigma$$

where:

- tensile stress, τ, is defined as: (force applied)/(cross-sectional area of the sample) $\tau = F/A$
- tensile strain, σ, is defined as: (extension)/(original length of the sample) $\sigma = \Delta x/L$
- and the constant of proportionality, E, is called the elastic modulus, or commonly Young's modulus.

Young's modulus is a measurable property of a given material, with a reliable value that can be used in calculations to predict the actual **strength** of components manufactured from it.

The relationships above are known as Hooke's Law. This is not a general law of physics, but is a very useful rule that is obeyed closely by many solid materials for a limited range of applied stresses – up to what is called their limit of proportionality.

For many materials, the limit of proportionality and the elastic limit may be quite close together, but they are not identical – the physically determined elastic limit and the mathematically derived limit of proportionality have different meanings and in general they will have different values.

Key terms

Plastic deformation – occurs under stress levels that are sufficient to make the solid material begin to flow, rather like a liquid. When the stress is removed, a change in an object's shape and size remains. This is called a **permanent set**.

Elastic limit – the point on the stress-strain curve, beyond which a material begins to suffer plastic deformation, and so will not completely regain its original dimensions when the stress is removed.

Strength – also called **ultimate tensile stress**, is the maximum stress that the material can bear. This occurs just before the material fails and fractures.

Stress-strain curves

Materials science involves a practical, experimental study of materials. Each material that has a slightly different chemical composition or manufacturing history – for example, a heat treatment or a shaping process – needs to have its tensile properties tested and recorded, so that its behaviour under load can be predicted and used in the design process. The results are most often presented graphically, as a stress-strain curve, and so you should become familiar with the shapes of these and how to interpret them.

Every stress-strain curve tells its own story about the internal structure of the material, including how tiny imperfections can grow or move under the influence of tensile stresses. These changes can lead to a hardening of the material, called work hardening, so that plastic deformation becomes much harder to produce. Their growth can lead to sudden brittle failure as cracks spread right across the width of the sample under test.

Ductile materials

The structure of solid metals is made up of tiny crystals – they are microcrystalline. The regular arrays of identical atoms make it relatively easy for layers of those atoms to slip across one another and find a new rest position that still keeps the crystal structure intact. This means that most metals show some **ductility** – i.e. under tensile stress they can be drawn out into a new shape, which they then take on permanently. Ductile materials can be drawn out into rods, wires or tubes, which makes them ideal for manufacturing.

Metals can generally also be formed by processes that use compressive force – hammering, rolling, stamping with a die. This is described by the term **malleability**. While all ductile materials are generally also malleable, there are some materials – e.g. lead – that are very malleable but which fail when put under quite small tensile stresses.

A lack of ductility or malleability is called **brittleness** – an example is glass.

Key terms

Ductility – the ability of a material to be formed by drawing into new shapes, primarily by means of tensile forces. Ductile materials are generally also malleable.

Malleability – the ability to be shaped by means of compressive forces such as occur in rolling, hammering or stamping. Not all malleable materials are also ductile.

Brittleness – the tendency of a material to fracture under stress. Brittle materials cannot undergo plastic flow and so are not suitable for manufacturing processes such as drawing, rolling, hammering or stamping.

Stress-strain curves for ductile materials have the basic characteristics shown in Figure 5.32. The stress values and the extents of strain vary markedly between different materials.

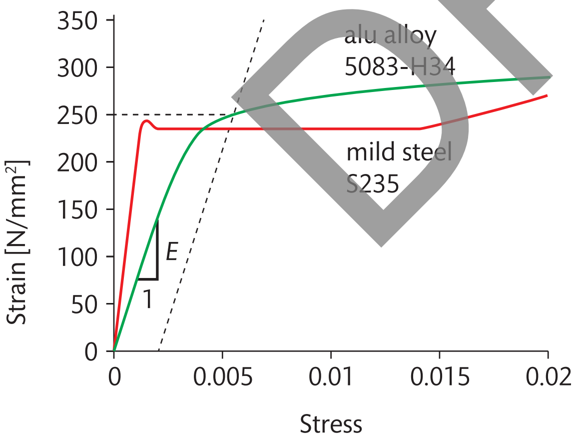


Figure 5.32 Stress-strain curves for an aluminium alloy and a mild steel

Creating alloys – mixtures of metals of varying compositions – is an important way of obtaining materials with properties suitable for specific applications. Aluminium and iron, perhaps the most widely used constructional metals, are almost never used in a pure form, but always alloyed with other elements. In iron (and steel, which is a class of iron alloys), the proportion of carbon included is especially important to both its strength and its ductility.

Iron and steel, along with just a few other metals, have a particular crystal structure that allows slippage to occur along planes in several different directions. The special ductility which this gives them is demonstrated on the stress-strain curve by a marked drop in stress as the **yield point** is passed. This important property links with iron and steel exhibiting a degree of protection against **fatigue** failure (see below) that other metals like aluminium do not share.

Key terms

Yield point – the point where the start of plastic flow causes a change of slope on the stress-strain curve. Iron and steel and a few other metals show a clearly defined yield with a drop of stress, while in many other materials the exact position of the yield point is hard to spot.

Fatigue – the embrittlement and failure of a material that can occur with relatively low levels of stress if these are repeatedly applied and then relaxed over many cycles.

Creep and fatigue

In a simple tensile test, the stress is gradually increased and the strain is recorded, generating the familiar stress-strain curve. It is a process that takes just a few minutes and is done once. But when materials are used in practical situations, the stress pattern is not always like that.

Typically, a component may have to carry a sustained stress for a long period of time. Design engineers use the yield stress as an indication of the strength of a material in tension. But a component that carries a stress well below its yield strength for a long period may begin to show a very slow deformation, which is called **creep**. This phenomenon becomes much more prevalent when components are working at a high temperature, even though still below their melting point. Random thermal energy of the atoms' microscopic vibrations combines with the applied stress to cause slippage and results in a plastic deformation.

Key term

Creep – sometimes called 'cold flow', occurs when a material under stress deforms gradually over time. It is more severe in materials that are subjected to heat for long periods.

Alternatively, some components have to withstand regular cycles of stresses increasing and then relaxing, only to increase again. Any part that is subject to mechanical vibrations will be affected in this way. Once again, the maximum stress allowed for in the component's design, including the extra stress due to the vibrations, may be well below the yield strength of the material. Nevertheless, these components often suffer catastrophic breakages known as fatigue failure.

The mechanism of fatigue at a microscopic level is a gradual growth of initially very small cracks. If the structure of the material were completely uniform, then the stress distribution across a component would also be smooth and even. But in reality the tiny crystals making up a metal object are of all sorts of shapes and sizes, and lie in random orientations. So the boundaries between them contain many irregularities. Impurities in a material add to the lack of regularity, as do tiny cracks that can occur due to contraction during rapid cooling and solidification. These imperfections mean that the internal distribution of stress is far from even, and stress concentrations occur, particularly around the tips of cracks. These potentially strength-limiting cracks can be seen if a component is examined using X-rays. The sharper the crack, the greater the stress concentration.

Under a cycling stress pattern, cracks grow progressively, and as they do the stress concentrations around them also increase. Finally, a critical point is reached where the concentrated stress is sufficient to cause a sudden brittle failure. In other words, a material that was originally strong and ductile has changed its character because of the long-term cyclic stress pattern that it has endured.

Case study

An example of where creep can cause a critical failure is in the vanes of a gas turbine engine. The high centrifugal forces and extremely high temperatures in parts of the engine could result in the blades gradually extending with creep until eventually they touch the engine casing at high velocity and rupture. But, before that happens, multiple creep-induced slippages along grain boundaries tend to form tiny voids, which initiate cracks. Cracks, once started, grow until failure occurs. The mechanism of the crack growth could involve a combination of further creep with some fatigue due to stress cycling as engine turbines speed up and slow down.

Check your understanding

1. Why are there high centrifugal forces in a gas turbine engine?
2. Why do creep-induced slippages happen along grain boundaries?

For most metals, if the cycling continues long enough, even quite low stress levels compared with the ultimate tensile strength will still eventually cause fatigue failure. So fatigue testing is carried out, where a component is deliberately tested under a cycling stress to determine a safe lifetime after which it must be replaced. This is an important part of the engineering design process, especially in critical applications like aeronautics.

Iron, steel and titanium alloys have the unusual property of having a well-defined fatigue limit – a stress level below which they can endure any number of cycles without developing fatigue failure. That means that a well-engineered component made in these materials can have an unlimited safe working life.

Theory into practice

In some cases ductility and malleability are wanted – mostly to aid in manufacturing items to particular shapes. For instance, metal drink cans are deep drawn from very ductile flat sheets of aluminium alloy. But in many other cases, resistance to deformation – hardness – is what is wanted – e.g. in the blades of knives and the tips of tools. The hardness of a metal can be increased by working it – stretching, hammering etc. – or by putting it through a heat treatment with rapid heating and cooling. Conversely, hardened objects can be softened again by slower cycles of heat treatment.

Reflect

Diamond is one of the hardest of materials, so one practical test for hardness uses a diamond cone and presses it into the surface of a test piece of material to see what size of impression it will make with a given force. Hardening increases the ultimate tensile strength of a metal, but it also makes it more liable to brittle failure.

In what laboratory applications might you want a hard material?

Elastomers

Objects made from rubber or synthetic polymers with rubber-like properties are able to exhibit large amounts of strain for a relatively small level of stress, but they still have the elastic property of springing back to their original dimensions. The elastic properties of metals are basically due to the strong chemical bonds between adjacent atoms being stretched under the applied stress, but having the ability to pull back afterwards. But the stretchiness of rubber is due to the very long molecules from which it is made. Under an applied stress, these can be stretched out nearly straight; but when the stress is removed, thermal motion quickly curls them back up into much shorter loops and knots.

If stress is applied gradually, and the material is warm, the same thermal motion can lead to the unravelling of some of the knots, so that a larger extension occurs. Afterwards, when the applied stress is removed, there may be a degree of permanent set. But if time is allowed for recovery, the random motion of the molecules gradually shortens and knots them up again, so recovering all or most of the original shape.

So, as a result, the stress-strain relationship for elastomer materials is heavily dependent on the rate at which the strain is induced. Moreover, the stress-strain curve for a reducing stress level follows a different path from that for a rising stress. This is called **elastic hysteresis**, and the shape traced out on the stress-strain graph is called a hysteresis loop (see Figure 5.33). The area enclosed by the hysteresis loop represents an amount of work done in stretching the material which is converted into thermal energy, and so is not recoverable as work in the relaxation phase of the cycle.

Key term

Elastic hysteresis – occurs in materials like rubber, where internal friction between large molecules dissipates energy producing heat. Loading and unloading of a sample each produces a different stress-strain curve, creating a hysteresis loop, the area of which represents the energy absorbed in the cycle.

The ability of rubber and materials like it to absorb mechanical energy and transform it into heat is very useful, and is why rubber is widely used for absorbing mechanical shocks and for damping vibrations.

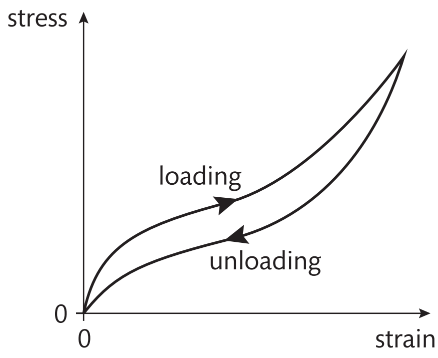


Figure 5.33 Stress-strain for rubber, showing a hysteresis loop. The loading and unloading curves are different.

Brittle materials

The majority of non-metallic solids are brittle in nature, the main exception being some of the long-chain polymer solids known as 'plastics'. All solids, including metals, become brittle if their temperature is lowered sufficiently to prevent any plastic flow occurring.

Brittle materials still have an elastic region in which applied stress produces a small strain, from which the material springs back when the stress stops. But they tend to be very stiff – i.e. they have large values for Young's modulus – and so stress rises quite sharply with only a small amount of strain (see Figure 5.34).

The microscopic structure of these materials doesn't allow for any plastic flow. Some of them are glassy, meaning that their atoms are rather randomly packed together with no long-range ordering – no crystalline regularity. Others, like concrete, may be composites made up of hard, strong particles held together by a binding medium. Composites that contain fibres, like wood, carbon fibre or fibreglass, may have a good tensile strength, but those like concrete that contain just granular material tend to be strong in compression, but very weak in tension. Concrete structures must either hold together by their weight, causing constant compression, or be reinforced with steel to take any possible tension loads.

The lack of any capacity for plastic deformation means that, as stress rises, the next event is a brittle failure. The stress concentration around the tips of small internal cracks or surface scratches in the material reaches a critical level, and the cracks propagate rapidly right across the structure, causing a sudden failure.

Some brittle materials can be very strong, but the lack of any yield point to give a warning of impending failure

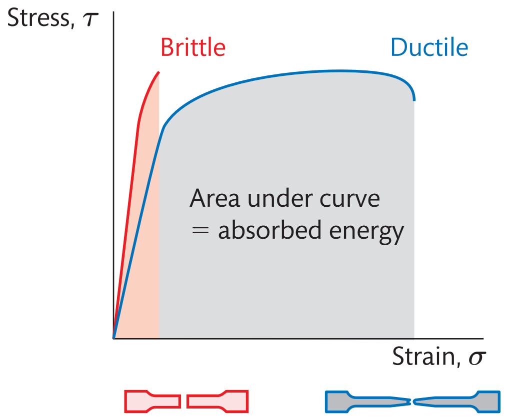


Figure 5.34 Comparing stress-strain relationships for brittle and ductile materials

means that loading on them needs to be carefully calculated and controlled. The high strength of glass in fibreglass stems from the fact that the glass fibres used are coated with a polymer resin immediately after manufacture and before they develop any surface scratches. It is those scratches that quickly weaken normal unprotected glass.

PAUSE POINT Close the book and try writing down what you now know about measuring and classifying the properties of materials.

Hint Think about materials you use at work and at home. Which need to be strong, hard, flexible, resilient . . . ? Which properties are particularly important for each?

Extend Look up in reference books or on the internet some tables of properties and some stress-strain curves for metals and non-metals. Which are the strongest, the stiffest, the most ductile?

Assessment practice 5.14

a) Sketch the tensile stress-strain relationships for a brittle material and for a ductile material, marking and labelling key features.

b) Diamond and steel are both strong materials. Referring to your sketches, explain why they are suitable for use in very different applications.

Assessment practice 5.15

a) State the meanings of yield strength and ultimate tensile strength.

b) Explain why engineers designing metal structures usually focus on the yield strength.

c) Creep and fatigue can both lead to failure at stress levels well below the yield strength. Describe under what conditions each of these may occur.

Fluids in motion

What makes fluids (i.e. liquids and gases) useful in systems is their ability to move – i.e. to flow, and in doing so to transmit pressure, to transfer heat or to simply deliver quantities of substance to a new location.

Fluid flow patterns – streamline and turbulent

Fluid flow fundamentally involves layers of molecules sliding over one another. For slow, gentle flow, motion occurs just in the direction of the applied stress, and is called streamline or laminar. This ordered motion does absorb some energy because of the viscous drag – a kind of internal friction – between the layers, but it is the most energy efficient kind of flow.

In streamline flow, the lines of flow are all parallel, and they can easily be made visible by introducing a dye, or by watching the motion of particles carried along by the liquid. Where the liquid is in contact with a solid surface, the speed of that layer is very close to that of the solid – stationary if it is the bank of a river, or moving if it is the hull of a boat or one of the rotating surfaces of an engine bearing. As you move through the liquid, across the direction of flow, the speeds of the layers progressively change. So, for example, in a river the fastest flow streams are found in the middle, well away from the banks. This means there is a velocity gradient perpendicular to the streams of flow.

Turbulent flow begins to appear as the velocity gradient is increased and the viscous drag forces also grow. The drag forces have a natural tendency to rotate parts of the fluid, and any obstacle or sharp corner that interrupts the smooth flow can set off turbulence, which will then spread to neighbouring parts. The rotational motion of turbulent flow absorbs much more energy, and so the effective drag force also becomes much larger. For that reason, except where stirring and mixing is needed, you will normally want to avoid turbulence as much as possible and to design for streamline flow.

Viscosity and Newton's Law

Dependent on their molecular make-up, some liquids offer much more viscous resistance to flow than others. So it is useful to have a measure of viscosity to compare them. This behaviour was first studied by Sir Isaac Newton, who defined a coefficient of dynamic viscosity, η, which is fundamental to exploring fluid flow.

Because the sliding of the layers of liquid over one another generates a resistance force, the kind of stress involved is a shear stress, equal to the size of that viscous drag force divided by the area of contact between the sliding layers. The 'strain' in this dynamic process is not a change in length or a distance moved, but rather the rate of shear strain, which is another name for the velocity gradient across the streams of flow.

Using those concepts, Newton's definition of viscosity is:

$$\text{coefficient of dynamic viscosity, } \eta = \frac{\text{shear stress, } \tau}{\text{rate of shear strain}}$$

$$\tau = \frac{F}{A} = \frac{\eta \ \Delta u}{\Delta y}$$

where:

- F is the viscous drag force generated across a layer of fluid of area A
- u is fluid velocity (in direction x) and y is distance perpendicular to the flow, so that $\Delta u/\Delta y$ is the velocity gradient or rate of shear strain (see Figure 5.35).

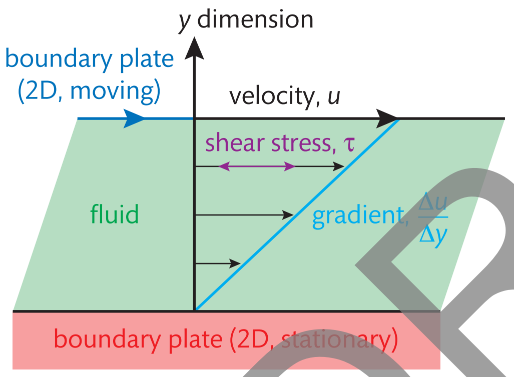


Figure 5.35 Defining viscosity

For the majority of simple fluids, the dynamic viscosity has a constant value at any given temperature. These liquids or gases are said to be Newtonian fluids, or to 'obey Newton's law of fluid flow'. As with Hooke's law for solids, Newton's law of fluid flow is not a general law of physics, but a useful rule obeyed by many fluids under streamline flow conditions.

An increase in temperature will generally lower the viscosity of a liquid, as thermal energy assists molecules in sliding over one another. But for gases, extra thermal energy increases the rate of collisions between the molecules, so gas viscosities increase when there is a rise temperature. You might notice this if you play badminton on a warm day.

Theory into practice

Terminal velocity: When an object falls under gravity through a viscous fluid – liquid or gas – it will reach a top speed, known as its terminal velocity. George Stokes studied this for small spherical objects falling relatively slowly so that the fluid flow around them was streamlined and he developed an equation, known as Stokes' Law, for the terminal velocity, v:

$$v = 2g\,r^2\,(\rho_1 - \rho_2)/9\,\eta$$

where g is the acceleration due to gravity, r is the sphere radius, ρ_1 and ρ_2 are the densities of the sphere and of the fluid medium, and η is the viscosity of the fluid.

This gives a simple practical method for measuring and comparing viscosity – simply dropping steel balls of known diameter into a long tube containing a fluid and measuring the terminal velocity. The equation is also useful when small droplets of liquid are falling through a gas (e.g. rain) or when small bubbles rise through a liquid.

Where more complex shapes and/or higher velocities result in turbulent flow – e.g. around a car, will the viscous drag be higher or lower than for streamline flow? What would this mean for the terminal velocity?

Investigation 5.2

Why not use Stokes Law and the steel ball-bearing method to devise your own investigation into how viscosities vary – e.g. a) between different kinds of oil; or b) as the temperature of a liquid is varied. For b) you would need to think about how you could ensure all the liquid was at the right temperature during each measurement.

Non-Newtonian fluid flow

Many rather more complex fluids have viscosities that change, either immediately or gradually, when they are sheared, e.g. by stirring or shaking. So these are called non-Newtonian. They are generally composed of large molecules in solution or of small solid particles or immiscible droplets of liquid or even gas bubbles in a liquid suspension. The interactions between the carrier liquid and what it contains result in the unusual rheological (i.e. viscous flow) properties of the combination fluid, and these can have various useful applications.

Non-Newtonian fluids may be classified into a number of types, as shown in Table 5.12.

Table 5.12 Non-Newtonian fluids

Fluid	Description	Examples
Shear thickening or 'dilatant' liquids	*Viscosity increases* immediately with a rise in the rate of shear strain	• Corn flour mixed with a small amount of water • Damp sand that firms when you walk on it
Shear thinning or 'pseudoplastic' liquids	*Viscosity decreases* immediately with a rise in the rate of shear strain	• Nail varnish • Most modern wall paints that brush out and then immediately set before they dry • Whipped cream • Quicksand – i.e. sand containing a higher proportion of water, which gives way and continues to thin as a victim thrashes around • Blood, which flows effectively under pressure through the vessels, but immediately thickens if it seeps out (a process that is followed by full coagulation, aided by platelet changes and clotting factor)
Rheopectic liquids	Gradually become *more viscous over time* when shaken, agitated, or otherwise stressed	• Synovial fluid, which provides both lubrication and shock protection in body joints by increasing its viscosity as it is being stressed • Some printer inks
Thixotropic liquids	Gradually become less viscous over time when shaken, agitated, or otherwise stressed	• Yoghurt • Ketchup that contains xanthan gum • Gelatin or pectin gels • 'One coat non-drip' paints that require a light brushing action and set fairly quickly, but not immediately, to give a single smooth, thick coating
Bingham plastics	Behave as a solid at low stresses but flow as a viscous fluid at higher stresses	• Toothpaste and similar pastes that extrude from a tube as a plug • Mayonnaise • Mud and slurries, which are important in drilling technology

Rate of fluid flow and pressure

A useful result for streamline flow of a Newtonian liquid down a pipe is Poiseuille's equation:

$$\frac{\Delta V}{\Delta t} = \frac{\pi r^4}{8\eta} \frac{\Delta p}{\Delta L}$$

That is, volume flow rate $\Delta V/\Delta t$ is directly proportional to the pressure gradient across the length of the pipe, $\Delta p/\Delta L$. The flow rate is very heavily dependent on the pipe internal radius, r, varying with the fourth power of r.

Exam tip

You don't need to learn this equation, but you should remember the meanings of 'flow rate', 'pressure drop' and 'pressure gradient', and have a practical grasp of how they work together.

If the flow becomes turbulent, the same equation can be used, but the viscosity, η, appears to increase. This is not a real change in viscosity, but just the effect of the extra energy absorbed because of the turbulence.

Gases are compressible, so the volume of a gas entering the pipe will be smaller than the volume on exit, where the measurement of volume flow rate is usually made. Nevertheless, the same formula for flow rate can be applied for gases if a correction factor is applied, multiplying it by (average pressure/outlet pressure).

So this formula can have wide application – including in medical fields such as the study of blood flow rates and air flow in lung alveoli, or applied to flow through a hypodermic needle.

Mass flow rate continuity

A simple but important observation for piped flow systems is that 'what goes in must come out'. This is true for mass flow rate across any pair of boundaries when a system is in steady flow. It can help in analysing what happens at pipe junctions where there are alternative flow paths, and it means that pressure drops and flow rates can be calculated for a complex network of pipes with different flow resistances using similar rules to those used for voltage, current and resistance in electrical circuits.

Bernoulli’s principle

Another important result comes from applying the principle of conservation of energy to fluid flow. It is called Bernoulli’s principle, and can be stated as follows:

> At any point in a tube through which fluid is flowing: the sum of the pressure energy, the potential energy and the kinetic energy is constant.

In fact, Bernoulli’s principle also applies to any smoothly flowing fluid, whether in a tube or not, and it has some very important applications.

Aerofoil lift

Aircraft wings, helicopter blades and sails for yachts are all designed with a curved profile that splits the airflow around them and creates two streams of flow (see Figure 5.36) with different lengths and hence different speeds. The faster air, travelling the longer distance, cannot exert as much pressure as the slower air-stream. So there is a net pressure difference across the aerofoil that gives it a ‘lift’ force. (Note that the ‘lift’ created by a curved sail on a sailing boat is a horizontal force, pulling the boat forward.)

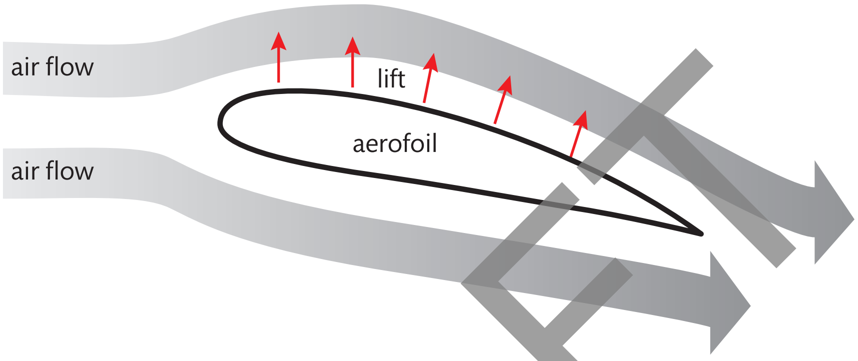


Figure 5.36 Aerofoil lift is created by splitting an airstream so that it travels further, and therefore faster, on one side than on the other.

Pitot tube velocity meters

Commonly used to measure the air speed of aircraft, but also in laboratory flow meters, a Pitot tube (see Figure 5.37) compares the pressure exerted by the moving fluid with a static pressure. The difference can be interpreted as an indication of the fluid velocity, using Bernoulli’s equation.

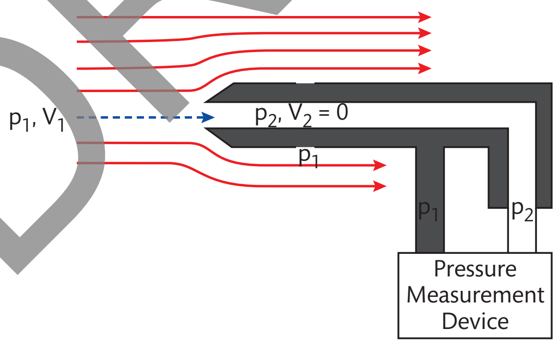


Figure 5.37 A Pitot tube velocity sensor senses the pressure of static air in its central tube, but also the lower pressure of moving air through annular holes.

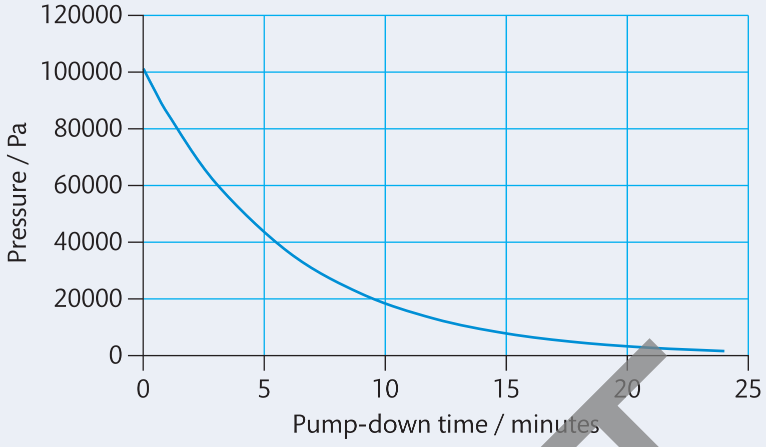
PAUSE POINT Close the book and make some notes on what you now understand about viscosity and fluid flow. What applications can you think of in your laboratory or your home?

Hint What laboratory equipment do you use to move fluids? – pipettes, syringes, vacuum pumps . . . How does the size of pipe or tube used affect the flow?

Extend Look around your kitchen or in your medicine or cosmetics cupboards for examples of non-Newtonian fluids. Test how they react to stirring, squeezing, shaking etc. and try to classify them according to type.

Assessment practice 5.16

The graph shows air pressure against time in a bell jar that is being evacuated by a rotary vacuum pump.



a) Explain the shape of the curve.

b) State which will improve the pump-down speed most: (i) halving the length of the connecting pipe, or (ii) doubling the internal diameter of the connecting pipe. Explain why.

Assessment practice 5.17

a) Describe the difference between streamline and turbulent flow patterns.

b) Give an example where turbulence is beneficial.

c) Explain why in most design applications streamline flow is preferred.

THINK FUTURE

Cemile Alkis

Research and development scientist in a medical laboratory

I started work at 22 just after I had left college. I worked in a chemical plant cataloguing materials and ensuring all the reactants we needed were of a high quality. I then moved jobs to work in a medical lab that develops and makes hip replacements. When I first started, I did a similar range of jobs to that in my first company, from cataloguing resources to testing materials. I have now worked with the company for five years and after two promotions I am the buyer for the company and I work with a range of organisations that produce our raw materials.

I spend a lot of time at their factory ensuring the quality of the production methods they use. I use my knowledge of extraction and purification techniques when doing this. My knowledge of the extraction process for titanium has been invaluable.

The scientists at my company give me a brief for the amount and quality of metal they need. I sometimes have to source raw materials from a range of companies.

I have to have good communication skills as I need to talk to everyone at the factory from the sales department to workers in the plant. I also need good maths skills as I have to negotiate good deals for my company. I do quite a lot of driving between my lab and the factories. I have to write month end reports to show what deals I have made and how I have met targets.

I often have to use online video conferencing to talk to my boss or people in the lab or factory. I nearly always have my laptop with me.

No two days are the same for me, which I really like. I get to use my science skills but also my other, personal skills. This challenges me and keeps the job interesting.

Focusing your skills

Think about the role of a buyer. Consider the following:

- What types of people will you work with and how will you support them?
- What role will you play in helping them achieve their goals?
- What experience would you need to carry out this role?
- What skills do you currently have? What skills do you think may need further development?

Getting ready for assessment

This section has been written to help you to do your best when you take the assessment. Read through it carefully and ask your tutor if there is anything you are still not sure about.

About the test

The test is in three sections (Biology, Chemistry and Physics).

Remember that all the questions are compulsory and you should attempt to answer each one. Consider the question fully and remember to use the key words to describe, explain and analyse. For longer questions, you will need to include a number of explanations in your response; plan your answer and write in detail.

Preparing for the test

To improve your chances on the test, you will need to make sure you have revised all the key assessment outcomes that are likely to appear. The assessment outcomes were introduced to you at the start of this unit.

To help plan your revision, it is very useful to know what type of learner you are. Which of the following sounds like it would be most helpful to you?

Type of learner	**Visual**	**Auditory**	**Kinaesthetic**
What it means	Need to see something or picture it, to learn it	Need to hear something to learn it	Learn better when physical activity is involved – learn by doing
How it can help prepare for the test	• Colour code information on your notes • Make short flash cards (so you can picture the notes) • Use diagrams, mind-maps and flowcharts • Use post-it notes to leave visible reminders for yourself	• Read information aloud, then repeat it in your own words • Use word games or mnemonics to help • Use different ways of saying things – different stresses or voices for different things • Record short revision notes to listen to on your phone or computer	• Revise your notes while walking – use different locations for different subjects • Try and connect actions with particular parts of a sequence you need to learn • Record your notes and listen to them while doing chores, exercising, etc. and associate the tasks with the learning

Remember!

Do not start revision too late! Cramming information is very stressful and does not work.

Useful tips

- **Plan a revision timetable** – schedule each topic you need to revise and try and spend a small time more often on each of them. Coming back to each topic several times will help you to reinforce the key facts in your memory.
- **Take regular breaks** – short bursts of 30–40 minutes revision are more effective than long hours. Remember that most people's concentration lapses after an hour and they need a break.
- **Allow yourself rest** – do not fill all your time with revision. You could schedule one evening off a week, or book in a 'revision holiday' of a few days.
- **Take care of yourself** – stay healthy, rested and eating properly. This will help you to perform at your best. The less stressed you are, the easier you will find it to learn.

Sitting the test

Listen to, and read carefully, any instructions you are given. Lots of marks are often lost because people do not read questions properly and then do not complete their answers correctly.

Most questions contain command words. Understanding what these words mean will help you understand what the question is asking you to do. These were also introduced at the start of this unit.

Remember the number of marks can relate to the number of answers you may be expected to give. If a question asks for two examples, do not only give one! Similarly, do not offer more information than the question needs: if there are two marks for two examples, do not give four examples.

Planning your time is an important part of succeeding on a test. Work out what you need to answer and then organise your time. You should spend more time on longer questions. Set yourself a timetable for working through the test and then stick to it. Do not spend ages on a short 1 or 2 mark question and then find you only have a few minutes for a longer 4 or 6 mark questions. It is useful when reading through a question to write down notes on a blank page. This way you can write down all the key words and information required and use these to structure an answer.

If you are writing an answer to a longer question, try and plan your answers before you start writing. Have a clear idea of the point your answer is making, and then make sure this point comes across in everything you write, so that it is all focused on answering the question you have been set.

If you finish early, use the time to re-read your answers and make any corrections. This could really help make your answers even better and could make a big difference in your final mark.

Hints and tips for tests

- Revise all the key areas likely to be covered. Draw up a checklist to make sure you do not forget anything!
- Know the time of the test and arrive early and prepared.
- Ensure that you have eaten before the test and that you feel relaxed and fresh.
- Read each question carefully before you answer it to make sure you understand what you have to do.

- Make notes as you read through the question and use these to structure your answer.
- Try answering all the simpler questions first then come back to the harder questions. This should give you more time for the harder questions.
- Remember you cannot lose marks for a wrong answer, but you cannot gain any marks for a blank space!

Q. Describe how transition metals are able to form ions with different oxidation numbers. (3)

Transition metals can lose two electrons from the 4s-orbital giving the +2 oxidation state. The 4s electrons are in the highest energy level and so are lost first. The 3d and 4s energy levels have similar energies so 3d electrons can also be lost.

This is a 3-mark describe question. The examiner is looking for 3 points. This answer has given 4 points, all of which are correct and so gains the marks. The examiner will not negatively mark the question but if you write more information that is incorrect and disagrees with what you have written you may lose a mark. For example, if this learner had gone on to write that 4s energy levels had much higher energy than 3d energy levels, this would be wrong and would negate the last mark given.

Q. Ethane contains a single C—C bond. Ethene contains a C═C double bond. Compare the bonds between the carbons in ethane and ethene. (6)

This is a 6-mark levelled question. It is worth 2 pass marks, 2 merit marks and 2 distinction marks. You gain marks for showing understanding rather than there being one mark per point. The more detailed and in-depth your discussion, the more likely you are to gain 6 marks. You would be expected to use all your knowledge of single and double bonds. You should discuss the electrons and orbitals, as well as strength, length and reactivity.

Question number	**Answer**	**Mark**
	Indicative content Similarities ▶ Both covalent bonds ▶ Electrons in the bonds overlap ▶ Both contain σ-bonds ▶ Hybrid sp^3 orbitals (ethane has sp^2) Differences ▶ Double bond contains π-bond ▶ *p*-orbitals overlap above and below the carbon atoms ▶ Movement restricted around double bond ▶ Region around double bond is flat ▶ Double bond more reactive than single bond/single bond more stable than double bond ▶ Single bond stronger than double ▶ Double bond shorter than single	(6)
0	No rewardable content.	0
Pass level	Some simple statements that are not linked.	1–2
Merit level	Some similarities and differences.	3–4
Distinction level	Detailed similarities and difference.	5–6

Ans 1. Double bonds are strong and very reactive. They make ethene a flat shape.

This would be a pass-level answer. The candidate has given some differences in a simple way. He has not compared double bonds to single bonds. This is worth 2 marks.

Ans 2. Both single and double bonds are covalent. The single bond is formed by a hybridised sp3 bond. In ethene there is also a pi bond. Both bonds contain a sigma bond and then in the double bond the p orbitals overlap. This forms a stronger bond but a shorter one. There are more electrons in a double bond so it is more reactive.

This would be a distinction-level answer. The learner has given similarities and differences. The ideas are mostly quite detailed and are linked. The learner does not have to give all indicative content to gain 6 marks.