1 Measurement

1.1 Quantities

A **physical quantity** is a property of a phenomenon, body or substance quantifiable by measurement.

Base quantities are physical quantities that are the most fundamental and independent of each other. **Derived quantities** are quantities formed by combining base quantities.

A **scalar** is a physical quantity with magnitude only. A **vector** is a physical quantity with both magnitude and direction.

A **homogeneous equation** is one in which every term has the same units. All physical equations are homogenous.

1.2 Error and uncertainty

Random error is an error in measurement in which measured quantities differ from the mean value by different magnitudes and directions. Sources of random error include parallax error, environmental variations, inherent irregularities, or equipment limitations (e.g. very sensitive equipment).

Systematic error is an error in measurement in which measured quantities differ from the true value by a fixed magnitude and in the same direction. Sources of systematic error include zero error, parallax error, and the environment (e.g. background radiation when measuring radioactive decay). A zero error occurs when an instrument registers a reading when there should be none.

Accuracy is a measure of how close the results of an experiment agree with the true value (systematic error). **Precision** is a measure of how close the results of an experiment agree with one another (random error).

1.2.1 Uncertainty propagation

For uncertainty propagation,

$$Y = ma \pm nb \implies \Delta Y = m\Delta a + n\Delta b \tag{1.1}$$

$$Y = a^m b^n \implies \frac{\Delta Y}{Y} = \left| m\frac{\Delta a}{a} \right| + \left| n\frac{\Delta b}{b} \right| \tag{1.2}$$

Base quantities and units

Quantity		Unit	
mass	m	kilogram	kg
length	l	metre	m
time	t	second	s
amount of substance	n	mole	mol
temperature	T	kelvin	K
current	I	ampere	A
luminous intensity	L	candela	cd

SI prefixes

Prefix	10^n	Prefix	10^n	Prefix	10^n
pico, p	-12	nano, n	-9	micro, μ	-6
deci, d	-1	kilo, k	3	mega, M	6

2 Kinematics

Distance x is the total length covered by a moving object irrespective of the direction of motion.

Displacement or position $\mathbf{x}$ is the shortest linear distance of a moving object from a given reference point.

Speed is the rate of change of distance travelled with respect to time.

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} \tag{2.1}$$

Velocity is the rate of change of displacement with respect to time.

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \tag{2.2}$$

Acceleration is the rate of change of velocity with respect to time.

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \tag{2.3}$$

Average speed is the total distance travelled over the total time taken.

$$\langle v \rangle = \frac{\Delta x}{\Delta t} \tag{2.4}$$

Average velocity is the total change in displacement over total time taken.

$$\langle \mathbf{v} \rangle = \frac{\Delta \mathbf{x}}{\Delta t} \tag{2.5}$$

Average acceleration is the total change in velocity over total time taken.

$$\langle \mathbf{a} \rangle = \frac{\Delta \mathbf{v}}{\Delta t} \tag{2.6}$$

2.1 Freefall

An object that experiences no force other than its weight and possibly drag is in freefall. Neglecting air resistance, an object in freefall near the Earth's surface experiences a constant downward acceleration $\mathbf{g} = 9.81\,\mathrm{m\,s^{-2}}$ towards the Earth's centre of mass.

In reality, an object will experience a drag force $\mathbf{F}_D \propto \mathbf{v}$ for small $\mathbf{v}$, and $\mathbf{F}_D \propto \mathbf{v}^2$ for larger $\mathbf{v}$. The drag force opposes the object's motion. Terminal velocity is the velocity an object reaches when drag exactly balances the object's weight.

2.2 Projectile motion

In **projectile motion**, it is assumed that **(a)** the acceleration due to gravity is constant throughout the motion, i.e. $\mathbf{g}$ is constant; **(b)** there is no horizontal acceleration, i.e. $\mathbf{a}_x = 0$; and **(c)** air resistance is negligible, i.e. $\mathbf{F}_D = 0$. Projectile motion with the above assumptions generally creates a parabolic trajectory that is symmetric; the **trajectory** is the path described by a projectile. **Range** is the distance on the plane between the point of projection and point of impact.

With air resistance, projectile motion describes an asymmetric trajectory. Air resistance also decreases the object's time of flight, horizontal range and maximum height reached.

2.3 Kinematics equations

The following equations apply only when acceleration is constant.

$$\mathbf{a} = \frac{\Delta \mathbf{v}}{\Delta t} = \frac{\mathbf{v} - \mathbf{u}}{t} \therefore \mathbf{v} - \mathbf{u} = \mathbf{a}t \therefore \mathbf{v} = \mathbf{u} + \mathbf{a}t \tag{2.7}$$

$$\mathbf{s} = \int \mathbf{v}\mathrm{d}t = \int (\mathbf{u} + \mathbf{a}t)\mathrm{d}t = \mathbf{u}t + \frac{1}{2}\mathbf{a}t^2 \tag{2.8}$$

$$\begin{aligned} \mathbf{s} &= \mathbf{u}t + \frac{1}{2}\mathbf{a}t^2 = \mathbf{u}t + \frac{1}{2}\frac{(\mathbf{v} - \mathbf{u})}{t}t^2 \\ &= \mathbf{u}t + \frac{1}{2}\mathbf{v}t - \frac{1}{2}\mathbf{u}t = \frac{1}{2}\mathbf{u}t + \frac{1}{2}\mathbf{v}t \\ &= \frac{1}{2}(\mathbf{v} + \mathbf{u})t \end{aligned} \tag{2.9}$$

$$\begin{aligned} \mathbf{v}^2 &= (\mathbf{u} + \mathbf{a}t)^2 = \mathbf{u}^2 + 2\mathbf{u}\mathbf{a}t + \mathbf{a}^2t^2 \\ &= \mathbf{u}^2 + 2\mathbf{a}(\mathbf{s} - \frac{1}{2}\mathbf{a}t^2) + \mathbf{a}^2t^2 \\ &= \mathbf{u}^2 + 2\mathbf{a}\mathbf{s} - \mathbf{a}^2\mathbf{t}^2 + \mathbf{a}^2\mathbf{t}^2 \\ &= \mathbf{u}^2 + 2\mathbf{a}\mathbf{s} \end{aligned} \tag{2.10}$$

3 Forces

3.1 Types of forces

Mass is the quantity of matter in an object, a measure of inertia, and the property of a body to resist changes in motion. As a result of mass, an object experiences **weight** $m\mathbf{g}$, which is the force acting on an object due to gravitational attraction. The **apparent weight** of an object is the normal contact force exerted by the surface on which the object is. The **centre of gravity** of an object is the point at which the entire weight of the object may be taken to act.

The **normal contact force** is the force exerted by the surface of one object on that of another when they are in physical contact; it acts perpendicularly to the surface of contact and prevents the objects from passing through each other. **Friction** is the resistive force acting between objects that opposes motion, preventing objects from slipping on each other. The **contact** or **reaction force** is the resultant of the friction and normal forces.

Drag is a resistive force experienced by an object moving in a fluid, which acts in the direction opposite to the object's motion.

Tension is the pulling force exerted by a rope, string, cable or other flexible object on another object.

3.2 Effects of forces

Hooke's law states that within the limit of proportionality, the extension produced in a material is directly proportional to the load applied. The tension created in a spring or material, within the constant of proportionality, is

$$\mathbf{F} = -k\mathbf{x} \tag{3.1}$$

The **moment** of a force is the product of the force and the perpendicular distance between the axis of rotation and the line of action of the force, given by

$$\boldsymbol{\tau} = \mathbf{F} \times \mathbf{d} = \mathbf{F}_{\perp}d \tag{3.2}$$

A **couple** is a pair of forces that are equal in magnitude but opposite in direction whose lines of action do not coincide. Their combined moment is the product of one of the forces and the perpendicular distance between the lines of action of the forces. A couple produces no resultant force as the forces balance. They produce only a resultant moment.

The **principle of moments** states that for an object to be in rotational equilibrium, the sum of clockwise moments about any pivot must equal the sum of anticlockwise moments about that pivot.

An object is stable if it returns to its original position after having been displaced slightly (by rotation). A low centre of gravity and a wide base helps to make an object more stable.

3.3 Pressure

Pressure is the force acting per unit area, where the force acts perpendicularly to the plane of the area A. The SI unit of pressure is pascal (Pa) i.e. $1\,\mathrm{N\,m^{-2}}$.

In a fluid with density ρ, the pressure acting on an object at depth h is

$$p = h\rho g \tag{3.3}$$

Take into account atmospheric pressure, or any other force acting on the liquid.

Upthrust is an upward force experienced by a body immersed in a fluid due to the pressure difference in the fluid between the lower and upper surface of the object.

$$\mathbf{U} = -V\rho\mathbf{g} \tag{3.4}$$

Archimedes' principle states that the upthrust experienced by an object partially or fully immersed in a fluid is equal to the weight of the fluid displaced by the object. For an object to float in a fluid, the maximum upthrust on the object if it is fully immersed in the fluid must be equal to or greater than the object's weight.

4 Dynamics

4.1 Newton's laws of motion

Newton's first law of motion states that a body will continue in its state of rest or uniform motion in a straight line unless an external resultant force acts on it.

Newton's second law of motion states that the rate of change of linear momentum of a body is directly proportional to the resultant force acting on it and the change takes place in the direction of the force.

Newton's third law of motion states that if object A exerts a force on object B, then object B exerts a force of equal magnitude but in the opposite direction on object A.

4.2 Momentum

The **linear momentum p** of an object is the product of its mass and velocity.

$$\mathbf{p} = m\mathbf{v} \tag{4.1}$$

The SI unit of momentum is N s. The **newton** (N) is the force required to effect a change in linear momentum at the rate of 1 kg m s^{-2}. The **impulse** of a force $\Delta\mathbf{p}$ is equal to the change in linear momentum caused by the force over some time period.

Average force is the impulse over total time taken.

$$\langle\mathbf{F}\rangle = \frac{\Delta\mathbf{p}}{\Delta t} \tag{4.2}$$

The **principle of conservation of momentum** states that the total momentum of a system of objects remains constant provided no external resultant force acts on the system.

In an **elastic collision**, both momentum and kinetic energy are conserved. In an **inelastic collision**, momentum is conserved but kinetic energy is not. In a **completely inelastic collision**, momentum is conserved but kinetic energy is not, and the objects stick together after collision.

For an elastic head-on collision of two objects, the relative speed of approach is equal to the relative speed of separation, regardless of the objects' masses.

$$\mathbf{v}_2 - \mathbf{v}_1 = \mathbf{u}_1 - \mathbf{u}_2 \tag{4.3}$$

5 Work, Energy and Power

The **work** done by a force on a body is the product of the force and the displacement of the body in the direction of the force.

$$W = \mathbf{F}\cdot\mathbf{s} = \left|\mathbf{F}_\parallel\right| |\mathbf{s}| \tag{5.1}$$

The SI unit of work done is the **joule** (J), which is the work done by a force of 1 N in moving an object by 1 m.

When a gas is expanding, it does work on external pressure of

$$W = p\mathbf{A}\cdot\mathbf{s} = p\Delta V = p(V_f - V_i) \tag{5.2}$$

When a gas is compressed, work is done on it by the surroundings, and work done by the gas is negative.

Energy is the stored ability to do work. Energy can be transferred through thermal transfer (heating), or mechanical transfer (by doing work).

There are many forms of energy, including kinetic energy, gravitational potential energy, elastic potential energy, chemical potential energy, electrical potential energy and light energy.

Kinetic energy is the energy possessed by a body due to its motion.

$$T = \frac{1}{2}m\mathbf{v}^2 \tag{5.3}$$

Derivation Consider a constant net force **F** acting on an object with mass m with some initial velocity **u**.

$$\mathbf{v}^2 = \mathbf{u}^2 + 2\mathbf{a}\mathbf{s}$$

$$\implies W = \mathbf{F}\cdot\mathbf{s} = m\mathbf{a}\mathbf{s}\ (\because \mathbf{F} = m\mathbf{a}) = \frac{1}{2}m(\mathbf{v}^2 - \mathbf{u}^2)$$

The work done by **F** increases only the kinetic energy of the block, so

$$W = \Delta T \therefore T_f - T_i = \frac{1}{2}m(\mathbf{v}^2 - \mathbf{u}^2)$$

If $\mathbf{u} = 0$,

$$T_i = 0 \therefore T_f = \frac{1}{2}m\mathbf{v}^2 \therefore T = \frac{1}{2}m\mathbf{v}^2 \qquad \blacksquare$$

Gravitational potential energy is the energy a body possesses due to its position relative to something else.

$$U = mgh \tag{5.4}$$

Derivation Consider a body of mass m at height h_i above the surface of the Earth, with acceleration of free fall **g**. A force **F** raises the body vertically at a constant velocity to height h_f.

$$\mathbf{a} = 0 \therefore \mathbf{F}_{\text{net}} = 0 \therefore |\mathbf{F}| - mg = 0 \therefore |\mathbf{F}| = mg$$
$$\therefore W = |\mathbf{F}|\,|\mathbf{s}| = mg(h_f - h_i)$$

The work done by **F** increases only the gravitational potential energy of the block, so

$$W = \Delta U \therefore U_f - U_i = mg(h_f - h_i)$$

If we take the gravitational potential energy at h_i as the reference level i.e. U at $h_i = 0$, then an object of mass m raised by height h vertically above the reference level has gravitational potential energy of $U = mgh$. $\blacksquare$

Elastic potential energy is the energy possessed by an elastic body when it is subjected to deformation.

$$E = \frac{1}{2}k(\Delta\mathbf{x})^2 \tag{5.5}$$

The principle of **conservation of energy** states that the total energy is constant i.e. energy cannot be created or destroyed, but can be converted from one form to another.

A **dissipative force** is a resistive force that reduces the energy of a system, like friction.

A non-isolated system is one where the system can interact and exchange energy with its environment. Energy of a non-isolated system is not conserved, but the energy of the system and its environment is always conserved.

Power is the work done per unit time, or the rate at which work is done.

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} \tag{5.6}$$

For a constant force $\mathbf{F}$ acting on an object such that the object moves with a constant velocity $\mathbf{v}$,

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{Fs}}{\mathrm{d}t} = \mathbf{F}\frac{\mathrm{d}s}{\mathrm{d}t} + \mathbf{s}\frac{\mathrm{d}F}{\mathrm{d}t} = \mathbf{F}\frac{\mathrm{d}s}{\mathrm{d}t} = \mathbf{F}\cdot\mathbf{v} \tag{5.7}$$

Efficiency is the ratio between the useful output of an energy conversion and the input.

6 Motion in a Circle

Angular displacement is the angle in radians through which a body is rotated about an axis.

$$\boldsymbol{\theta} = \frac{s}{r} \tag{6.1}$$

One **radian** is the angle subtended at the centre of a circle by an arc that is equal in length to the radius of the circle.

Angular velocity is the rate of change of angular displacement w.r.t. time.

$$\boldsymbol{\omega} = \frac{\mathrm{d}\boldsymbol{\theta}}{\mathrm{d}t} \tag{6.2}$$

The **period** of an object in circular motion is the time taken for the object to make one complete revolution, and the frequency of an object in circular motion is the number of complete revolutions made by the object per unit time.

$$T = f^{-1} = \frac{2\pi}{\omega} \tag{6.3}$$

$$f = T^{-1} = \frac{\omega}{2\pi} \tag{6.4}$$

The **linear velocity** of an object moving in a non-linear path is tangential to the object's path.

$$\mathbf{v} = \mathbf{r}\times\boldsymbol{\omega} \tag{6.5}$$

An object in **uniform circular motion** has constant angular velocity and a constant linear speed.

The **centripetal acceleration** (acceleration in the direction of the centre of motion) is

$$\mathbf{a}_c = \frac{\mathbf{v}^2}{\mathbf{r}} = \mathbf{r}\boldsymbol{\omega}^2 \tag{6.6}$$

For the centripetal acceleration to exist, the net force must be in that direction. A net force in the direction of the centre of motion is known as a centripetal force.

$$\mathbf{F}_c = m\frac{\mathbf{v}^2}{\mathbf{r}} = m\mathbf{r}\boldsymbol{\omega}^2 \tag{6.7}$$

In some cases, friction can be the force providing for the centripetal force.

7 Oscillations

Oscillation is the repetitive variation, typically in time, of some measure about a central value (often a point of equilibrium) or between two or more different states.

7.1 Simple harmonic motion

Simple harmonic motion is the oscillatory motion of a particle whose acceleration is always directed towards a fixed point and is directly proportional to its displacement from that fixed point i.e. it is the motion that satisfies the equation

$$\mathbf{a} = -\omega^2\mathbf{x} \tag{7.1}$$

The **period** T of an object in simple harmonic motion is the time it takes to make one complete oscillation. Likewise, the frequency f, with units Hz, is the number of complete oscillations per unit time.

Phase φ is the wave angle, or the fraction of, which has elapsed relative to some defined point, typically the last complete oscillation. It is usually given in radians, between 0 rad to 2π rad.

$$\varphi = 2\pi\frac{t}{T} \tag{7.2}$$

Phase difference is simply the difference in phase between two particles, or one particle at different times.

The **amplitude** x_0 is the maximum magnitude of displacement of the oscillating particle from equilibrium.

The displacement of a particle at a given time is

$$\begin{aligned}\mathbf{x}(t) &= x_0\sin(\omega t)\\ &= x_0\cos(\omega t)\end{aligned} \tag{7.3}$$

The velocity of a particle is given by

$$\begin{aligned}\mathbf{v}(t) &= \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \omega x_0\cos(\omega t)\\ &= -\omega x_0\sin(\omega t)\\ \mathbf{v}(\mathbf{x}) &= \pm\omega\sqrt{{x_0}^2 - \mathbf{x}^2}\end{aligned} \tag{7.4}$$

The acceleration of a particle is given by

$$\begin{aligned}\mathbf{a}(t) &= \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\omega^2 x_0\sin(\omega t)\\ &= -\omega^2 x_0\cos(\omega t)\\ \mathbf{a}(\mathbf{x}) &= -\omega^2 x\end{aligned} \tag{7.5}$$

Based on the above equations, it can be said that

$$\mathbf{x} = 0 \implies v = \max_{\mathbf{x}} v = \omega x_0 \vee \mathbf{a} = 0$$

$$\mathbf{x} = \pm x_0 \implies \mathbf{v} = 0 \vee a = \max_{\mathbf{x}} a = \omega^2 x_0$$

If the displacement an object in uniform circular motion is broken into two perpendicular axes coplanar with the plane of circular motion, the motion in both axes separately are simple harmonic.

Examples of simple harmonic motion include the simple pendulum for small angles, where

$$\omega^2 = \frac{g}{l} \tag{7.6}$$

and a mass on a spring for a spring obeying Hooke's law, where

$$\omega^2 = \frac{k}{m} \tag{7.7}$$

The energies of an object in simple harmonic motion are given by

$$\begin{aligned} T &= \frac{1}{2}m\omega^2{x_0}^2\cos^2(\omega t) \\ &= \frac{1}{2}m\omega^2{x_0}^2\sin^2(\omega t) \\ &= \frac{1}{2}m\omega^2({x_0}^2 - \mathbf{x}^2) \end{aligned} \tag{7.8}$$

If we assume the potential energy at equilibrium to be zero, then

$$\begin{aligned} U &= \frac{1}{2}m\omega^2{x_0}^2\sin^2(\omega t) \\ &= \frac{1}{2}m\omega^2{x_0}^2\cos^2(\omega t) \\ &= \frac{1}{2}m\omega^2\mathbf{x}^2 \end{aligned} \tag{7.9}$$

This gives us total energy

$$E = \frac{1}{2}m\omega^2{x_0}^2 \tag{7.10}$$

7.2 Damping

Dissipative forces like air resistance and friction typically oppose oscillating systems in real life. The amplitude and total mechanical energy of real systems generally decrease with time i.e. they are damped.

Examples of damped oscillations include a note played on a piano gradually fading away after being played, or a bell being loud when it is struck but gradually fades away after.

There are three degrees of damping.

- **Light damping** occurs when there are still oscillations, but the amplitude of oscillation decreases gradually over time.
- **Critical damping** occurs when the system returns to its equilibrium position in the shortest possible time without any oscillation i.e. it does not cross the equilibrium position at all.
- **Heavy damping** occurs when the system returns to its equilibrium position very slowly, without any oscillation. The system does not cross the equilibrium position either.

Critical damping is important and seen in many places in our daily lives. In instruments like balances and electric meters, the pointer is critically damped so that it moves quickly to the reading without oscillating. In car suspension systems, the shock absorbers critically damp the suspension and resist the setting up of vibrations that could impair control or cause damage.

7.3 Forced oscillations and resonance

A system undergoing free oscillation will oscillate with a frequency known as its **natural frequency** f_0.

A **forced oscillation** is the motion produced when a system (the driven oscillator) is acted upon by an external periodic force (the driving force). The system will oscillate at the driving frequency.

The amplitude of oscillation of a system depends on the driving frequency, and reaches a maximum when the driving frequency matches the natural frequency. When this occurs, it is known as **resonance**, and there is a maximum transfer of energy from the driving system into the oscillating system.

If the system is not damped, the amplitude of the driven system in resonance will tend towards infinity. Otherwise, the peak amplitude and natural frequency decreases as the degree of damping increases, as greater damping means a greater period and smaller frequency.

Resonance can be useful. It is exploited to produce sound in wind instruments and is also the basis of modern radio telecommunications, where electromagnetic resonance occurs between the radio circuit and the signal. It also allows for cooking of food using microwaves, where microwaves resonate with water molecules.

However, resonance can be harmful. It can cause damage to human organs – internal organs can resonate with sounds of frequency below 100 Hz. It can also cause bridges to collapse due to resonance – like with the wind e.g. the Tacoma Narrows Bridge, or when people march in step over bridges (which is why armies break step when crossing bridges).

8 Thermal Physics

The **internal energy** U of a system is the energy due to its constituent matter, and is determined by the state of the system. It is the sum of the microscopic kinetic energy of all the atoms or molecules due to their continuous random motion, and the microscopic potential energy of all the atoms or molecules due to their positions relative to each other and the forces between them.

Temperature T is a measure of the average translational kinetic energy of a system's atoms or molecules. The SI unit of temperature is the kelvin, K. When a body's temperature increases, its internal energy increases.

Objects are in **thermal contact** when heat can be exchanged between them. Objects in thermal contact are in **thermal equilibrium** when the net exchange of heat between the two is zero.

Heat Q is the thermal energy exchanged between two objects due to a difference in temperature. Heat always flows from an object of higher temperature to one of lower temperature. Thus, objects with the same temperature are in thermal equilibrium.

The **thermodynamic** (Kelvin) **scale** is an absolute scale of temperature that does not depend on the properties of any particular substance, although fixed points are defined based on water. One kelvin is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

$$T/\mathrm{K} = T/^{\circ}\mathrm{C} + 273.15 \tag{8.1}$$

On the thermodynamic scale, absolute zero (0 K) is the temperature at which all substances have a minimum internal energy.

An **empirical temperature scale** depends on experimental results while the thermodynamic scale is theoretical; the former depends on a thermometric property of some substance, but the thermodynamic scale is independent of the properties of any substance.

8.1 Heat capacity and latent heat

The **heat capacity** C of a body is the quantity of heat required to produce a unit change in temperature without a change in phase.

$$Q = C\Delta T \tag{8.2}$$

Heat capacity is an extensive property.

The **specific heat capacity** c of a body is the quantity of heat required to produce a unit change in temperature per unit mass of a substance without a change in phase.

$$Q = mc\Delta T \tag{8.3}$$

Specific heat capacity is an intensive property, and is a characteristic of the material of the body.

Heat capacity is not a constant, and is affected by the state of a body – p, V and T. Values of heat capacity are generally an average over the temperature range.

The heat capacity (and by extension, specific heat capacity) of a substance can be determined by electrical heating. Generally, the electrical energy supplied IVt is equated by conservation of energy to energy absorbed by the substance, and possibly heat lost to the environment.

Latent heat is energy required for a substance to change phase. The determination of latent heat by electrical methods follows the same principles as that of heat capacity.

The **specific latent heat of fusion** L_f is the quantity of heat required to change a unit mass of a substance from the solid phase to the liquid phase without any change in temperature.

The **specific latent heat of vaporisation** L_v is the quantity of heat required to change a unit mass of a substance from the liquid phase to the gaseous phase without any change in temperature.

During melting or boiling, heat is supplied to the system. Microscopic potential energy of the system increases while microscopic kinetic energy remains unchanged. Since microscopic kinetic energy is unchanged, the temperature of the system remains unchanged.

Molecules in a solid are bound by intermolecular forces and vibrate around their equilibrium position in a crystalline structure. When melting, there is no increase in temperature until the entire solid has melted; molecules gain energy to overcome the intermolecular forces that hold them together. They then acquire a greater degree of freedom and disorder that characterises the liquid phase. Latent heat of fusion is used to weaken the intermolecular attractive forces and increase the separation between molecules.

Molecules in a liquid are close together, as forces between molecules are stronger than in a gas, where the spacing between molecules is much further apart. When boiling, there is no increase in temperature until all of the liquid has boiled; molecules gain energy to overcome intermolecular forces that hold the molecules together and increase the separation between molecules. Some latent heat of vaporisation is used to allow the vapour to expand against atmospheric pressure.

The latent heat of vaporisation is always greater than the latent heat of fusion, as when melting, molecules need only to break down the structure into a less-ordered arrangement of molecules, but when vaporising, molecular bonds are nearly completely broken and this requires much more energy. Also, when melting, a substance's volume does not increase substantially, but when vaporising, the increase in volume is much greater, and so much more work needs to be done against atmospheric pressure when vaporising than melting, leading to latent heat of vaporisation being greater.

Evaporation is the change in state from liquid to gas at the surface of a liquid at any temperature. It occurs as molecules in a liquid are in continuous random motion and collide frequently with one another, causing the speed of a given molecule to change continually. If at the surface, a molecule moving away from the liquid has enough kinetic energy to do work against intermolecular attractive forces and atmospheric pressure, it can escape from the liquid and become a molecule of the vapour.

Cooling occurs during evaporation as when the more energetic molecules escape from the surface of the liquid, the remaining molecules will have a lower mean translational kinetic energy. The overall mean kinetic energy of the remaining molecules thus decreases, so the temperature decreases, so cooling occurs.

The rate of evaporation can be increased by increasing the temperature or surface area of the liquid, adding wind, or reducing the air pressure above the liquid.

8.2 State equation and the first law of thermodynamics

The **first law of thermodynamics** states that the internal energy of a system is a function of its state; an increase in internal energy of a system is equal to the sum of the heat supplied to the system and the work done on the system.

$$\Delta U = Q + W \tag{8.4}$$

One **mole** is the amount of substance that contains the same number of elementary entities as there are in 12 g of carbon-12. This number is known as **Avogadro's constant**, N_A.

The **ideal gas law** gives the equation

$$pV = nRT \tag{8.5}$$

An ideal gas is one where **(a)** particles are all moving randomly and Newton's laws of mechanics can be applied to an individual particle's motion; **(b)** all atoms or molecules are identical and spherical; **(c)** the total volume of the particles is negligible compared to the total volume of the container; **(d)** there are no forces of attraction or repulsion between the particles, except during a collision; **(e)** collisions between the particles are perfectly elastic; and **(f)** the time taken for a collision is negligible. These assumptions obviously do not apply to reality; all gases in reality are real gases.

The mean translational kinetic energy of a monatomic ideal gas particle is

$$T = \frac{1}{2}m\langle \mathbf{c}^2 \rangle = \frac{3}{2}kT \tag{8.6}$$

resulting in the total kinetic energy being

$$T = \frac{3}{2}NkT = \frac{3}{2}nRT \tag{8.7}$$

The internal energy of a monatomic ideal gas is equal to its kinetic energy – since there are no forces of attraction, there is no potential energy.

$$U = KE = \frac{3}{2}NkT = \frac{3}{2}nRT \tag{8.8}$$

Work done by an ideal gas during expansion is

$$W = p\Delta V = nR\Delta T = Nk\Delta T \tag{8.9}$$

There are generally four types of processes.

- Isochoric: volume is kept constant.

$$\Delta V = 0 \implies \Delta W = 0$$

- Isobaric: pressure is kept constant

$$\Delta p = 0$$

- Isothermal: temperature is kept constant.

$$\Delta T = 0 \implies \Delta U = 0$$

- Adiabatic: there is no heat gained or lost.

$$Q = 0$$

9 Wave Motion

A **wave** is a disturbance or oscillation that travels through space, accompanied by a transfer of energy, but without any transfer of material between the points.

Waves are mechanical or electromagnetic, transverse or longitudinal, and progressive or stationary.

A mechanical wave requires a medium for propagation, while electromagnetic waves do not. In mechanical waves, particles of the medium oscillate, while in electromagnetic waves, the electric and magnetic fields oscillate. In either case there is no propagation of material.

Transverse waves have particles oscillating perpendicularly to the direction of wave propagation, while **longitudinal waves** have particles oscillating parallel to the direction of wave propagation.

In a **progressive wave**, the profile of the wave appears to be moving, and energy is transferred. In a **standing wave**, the profile of the wave does not appear to move, and energy may or may not be transferred.

The **crest** on a sinusoidal waveform is the highest point while the **trough** is the lowest point. The **wavelength** λ is the distance between two successive points on a wave that are in phase.

The (phase) velocity of a wave is the velocity at which the phase of the wave travels.

Derivation If we assume it is constant, then velocity is the distance travelled by the wave in one cycle over the time taken for one cycle of the wave. For one cycle of the wave, the distance travelled is the wavelength λ and time taken is the period T. Thus the speed of the wave

$$v = \frac{\lambda}{T} = f\lambda \tag{9.1}$$

■

9.1 Graphs

The angular wavenumber (or simply wavenumber) k is the radians per unit distance of the wave.

$$k = \frac{2\pi}{\lambda}$$

The displacement-time graph of a single particle in a wave is given by

$$y = y_0 \sin(\frac{2\pi}{T}t) = y_0 \sin \omega t \tag{9.2}$$

It shows the displacement of a single particle along the path of propagation at various times.

The displacement-distance graph of a wave is given by

$$y = y_0 \sin(\frac{2\pi}{\lambda}x) = y_0 \sin k\lambda \tag{9.3}$$

It shows the displacement of all particles along the path of propagation at various distances from the source at an instant in time.

Wavefronts show the position of points of a wave that are in phase. The direction of wave propagation is perpendicular to the wavefronts.

For a longitudinal wave, if displacement to e.g. the left is taken as negative, and to the right as positive, then the displacement-time and displacement-distance graphs can also be plotted.

9.2 Phase

Phase and phase difference are as defined in oscillations. However, phase difference can also be determined using the path difference Δx:

$$\Delta\phi = \frac{\Delta x}{\lambda}2\pi \tag{9.4}$$

Particles in phase have a phase difference of 0 rad and a path difference that is an integer multiple of the wavelength. Particles in antiphase have a phase difference of π rad and a path difference that is an odd multiple of $\lambda/2$ i.e. $(n+\frac{1}{2})\lambda$, $n \in \mathbb{Z}_0^+$.

9.3 Intensity

The **intensity** of a wave is the rate of transfer of energy per unit area, with units $\mathrm{W\,m^{-2}}$. In three dimensions, intensity is inversely proportional to the square of distance; in two dimensions, it is inversely proportional to distance only, since the energy spreads out only on the circumference of a circle.

For a point source that spreads out uniformly from the source, energy is spread out in 3 dimensions on an expanding spherical surface. Since the area of a sphere is $4\pi r^2$, then

$$I = \frac{P}{4\pi r^2} \tag{9.5}$$

A directed source spreads out over the area of a hemisphere, and so on.

Intensity is always proportional to the square of amplitude, for both 2 and 3 dimensions i.e.

$$I \propto {x_0}^2 \tag{9.6}$$

9.4 Polarisation

Polarisation is a property of waves that can oscillate with more than one orientation i.e. for transverse waves. A wave is linearly polarised when it oscillates in a single plane only; it is circularly polarised when the plane of oscillation rotates with time, but waves only oscillate in that plane.

When referring to polarised electromagnetic waves, by convention, we refer to the electric field polarisation.

A polariser blocks waves that do not oscillate along the polarising axis. Electric field waves that are parallel to the polarising axis will be transmitted while those that are perpendicular to the polarising axis will be blocked. For oscillations not parallel or perpendicular to the polarising axis, they can be resolved into the two axes. Light that emerges from a polariser will be polarised along the polariser's polarising axis.

The intensity of polarised light after passing through a polariser

$$I = I_0 \cos^2\theta \tag{9.7}$$

where θ is the angle between the plane the light is polarised in, and the polariser's polarising axis.

Longitudinal waves cannot be polarised as the direction of oscillation of particles is already in a single plane (parallel to the direction of wave propagation), and they will not be restricted by the polarising axis.

9.5 Determination of f and λ

Sound waves create pressure differences in the air that can be detected and converted into electrical signals by a microphone, which is connected to the Y-input of a cathode ray oscilloscope to display the electrical signal. The time-base should be adjusted to a suitable sweep frequency so a sinusoidal trace can be displayed. The period of the trace can be read off after determining the scale of the time-base.

To determine the wavelength of a sound, use a reflector set perpendicularly to the direction the sound is coming from, which will form a stationary sound wave. Use a microphone connected to a cathode ray oscilloscope to find a point (a displacement node) where maximum amplitude is detected on the CRO. Mark this point. Then find the next adjacent point where there is maximum amplitude, and mark this point. Measure the distance between the two marked points and double it; this is the wavelength of the sound wave.

10 Superposition

Superposition is a property of linear systems where the net response at a given point and time caused by two or more stimuli is the sum of the responses which would have been caused by each stimulus individually.

When contextualised to waves, the **superposition principle** is as such: When two waves of the same kind meet at a point in space, the resultant displacement at that point is the vector sum of the displacements that the two waves would separately produce at that point.

10.1 Interference

Interference is the superposing of 2 or more waves of the same type, interacting according to the principle of superposition. **Observable interference** is the superposing of 2 or more coherent waves to produce regions of maxima and minima in space, according to the principle of superposition.

Coherent sources are sources with the same frequency and a constant phase difference.

For interference to be observable, the sources must be coherent; the waves must have approximately equal amplitudes, and they must be polarised in the same plane, or unpolarised.

Constructive interference occurs when two or more waves arrive at the screen in phase with each other, such that the amplitude of the resultant wave is the sum of the amplitudes of the resultant waves.

Destructive interference occurs when two or more waves arrive at the screen π out of phase with each other, such that the amplitude of the resultant wave is the minimum possible value.

Interfering waves produce interference patterns with alternating maxima and minima. The zero order maximum is the brightest maximum; if there are multiple maxima with the highest amplitude, the zero order can be chosen as any. Successive maxima away from the 0 order are numbered the 1st, 2nd, 3rd and so on. Minima have the same order as the adjacent maximum away from the 0 order. This means that there is no 0 order minimum.

10.1.1 One-source interference i.e. diffraction

Diffraction is the bending or spreading of waves when they travel through a small opening. It is a special case of interference, and it occurs due to the Huygens-Fresnel principle, where each point along a slit is considered a source of spherical waves.

In diffraction, waves from the Huygens sources interfere with each other and produce a pattern with alternating maxima and minima, with minima at angles θ satisfying the equation

$$a \sin\theta = m\lambda \tag{10.1}$$

where a is the slit width and $m \in \mathbb{Z}^+$ is the order of the minimum.

Generally, diffraction can only be observed when the size of the opening is approximately the same order as the wavelength or smaller.

10.1.2 Two-source interference

Two point sources of spherical waves can produce an interference pattern on a screen, with alternating maxima and minima. The zero order maximum occurs at the point on the screen that is equidistant from both sources (assuming the sources are in phase).

The separation between successive maxima is estimated by

$$\Delta x = \frac{\lambda D}{a} \tag{10.2}$$

for $D \gg a$, where D is the distance from the point sources to the screen, and a is the separation of the two point sources. If $D \approx a$ or $D < a$, then Pythagoras' theorem must be used.

This phenomenon is investigated in Young's double slit experiment, where the two point sources are replaced by a monochromatic source of light passing through a single slit that acts as a point source, followed by a double slit that acts as two coherent point sources and causes light to diffract so that they can overlap and interfere.

For such a setup to produce a visible pattern, the double slits must be small enough for diffraction to occur; the single slit must be small enough so the light reaching the double slits is coherent; and the slit-screen separation must be at least one order larger than the slit separation so that the diffracted waves can overlap and interfere.

The path difference Δx between the two waves reaching a maximum and minimum in two-source interference can be expressed in terms of the order of the maximum n or minimum n. If the two sources are in phase, then for maxima

$$\Delta x = n\lambda \tag{10.3}$$

and for minima

$$\Delta x = (n - \frac{1}{2})\lambda \tag{10.4}$$

and vice versa for sources in antiphase.

Two-source interference can be demonstrated with a ripple tank. Two dippers connected to a bar connected to a vibrator are set into vertical vibrations with the same amplitude and frequency and in phase. Each dipper produces circular waves that have the same frequency, which spread out and overlap, interfering constructively and destructively, and an interference pattern where there are points with water waves of maximum amplitude and points with minimum amplitude is formed.

Two-source interference can also be demonstrated with a microwave. Microwaves from a transmitter pass through two slits on an aluminium plate that act as coherent sources of microwaves. Microwaves emerging from the slits diffract and overlap, resulting in constructive and destructive interference. A microwave detector is moved along a line and the current registered will vary alternately from a maximum to a minimum, showing constructive and destructive interference. Since the wavelength of microwaves is about 3 cm, the slit size should be about 3 cm, and the slit separation should be about 30 cm, about 10λ. The interference can be observed about 1 m away.

10.1.3 Many-source interference

A **diffraction grating** is a plate with a large number of parallel, identically spaced slits of the same width. These multiple slits act as multiple coherent sources.

Multiple sources of waves can produce an interference pattern similar to the previous two, with alternating maxima and minima. The zero order beam is the one parallel to the original beam of light.

The angle θ from the normal to the grating at which each maximum (or each beam) occurs satisfies the equation

$$d \sin\theta = m\lambda \tag{10.5}$$

where d is the slit separation, and m is the order of the maximum.

The maximum observable order for a planar screen parallel to the grating occurs at

$$0 < \theta < 90^\circ \iff \sin\theta < 1$$

From the equation above, it can be derived that

$$m < \frac{d}{\lambda} \tag{10.6}$$

10.2 Standing waves

When two waves of the same type, amplitude and frequency travel in opposite directions to each other and overlap, a standing wave will be formed due to superposition.

Standing waves have points where constructive and destructive interference always occurs; they are known as **antinodes** and **nodes** respectively. Nodes are halfway between antinodes, and vice versa. Nodes have the minimum amplitude and intensity, while antinodes have the maximum amplitude and intensity.

In stationary waves, each point between consecutive nodes are in phase with each other i.e. they will reach the amplitude and equilibrium at the same time. Points to the between a node and the next node to the left are in antiphase with points between that node and the node to the right.

The distance between consecutive nodes or consecutive antinodes is half the wavelength of the original waves.

Standing waves can be produced when a reflector reflects waves back in the direction they came from. In reality, however, reflected waves will not have the same amplitude, so perfect standing waves are not formed.

10.2.1 Harmonics

A **harmonic** of a wave is a frequency that is an integer multiple of the fundamental frequency.

Standing waves can be produced with nodes at both ends of the wave. This occurs e.g. in a string fixed at both ends that is then plucked. The longest possible wavelength occurs when the two end nodes are consecutive nodes i.e. the distance L between the two nodes is half the wavelength. This mode is known as the 1st harmonic or the fundamental. The frequency, which is the fundamental frequency, is then

$$f_1 = \frac{v}{2L}$$

The next possible standing wave occurs when there is one node between the two end nodes i.e. the distance L is the wavelength. This mode is the 2nd harmonic or the 1st overtone and has frequency

$$f_2 = \frac{v}{L} = 2f_1$$

Therefore, the frequency and wavelength of the nth harmonic or $(n-1)$th overtone are

$$f_n = nf_1$$
$$\lambda_n = \frac{2L}{n}$$

Standing waves can be produced with a node at one end and an antinode at the other. This occurs in a sound wave that enters an air column with one end closed, for example. The longest possible wavelength occurs also when the end node and antinode are adjacent i.e.

$$L = \frac{1}{4}\lambda$$

Thus the 1st harmonic or fundamental

$$f_1 = \frac{v}{4L}$$

The next possible standing wave occurs when there is one node between the node and antinode i.e.

$$L = \frac{3}{4}\lambda$$

This is the *3rd* harmonic (since f is thrice that of f_1) or 1st overtone;

$$f_3 = 3\frac{v}{4L} = 3f_1$$

Therefore, the frequency and wavelength of the nth harmonic or $\frac{1}{2}(n-1)$th overtone are

$$f_n = nf_1$$
$$\lambda_n = \frac{4L}{n}$$

Standing waves can be produced with antinodes at both ends. This occurs in a sound wave that enters an air column open at both ends. The modes of operation for this are identical to standing waves with nodes as ends.

In general, closed ends are usually nodes as particles cannot vibrate there; conversely, open ends are antinodes as particles must vibrate there.

For sound waves, pressure nodes are displacement antinodes, and vice versa. At a pressure node, a microphone or ear will not detect sound as they sense pressure variations, not displacement.

The air at the end(s) of a pipe are generally free to move and so naturally vibrations will extend into the air outside the pipe i.e. the antinode is slightly beyond the open end. The extra length it extends is called the **end correction**, and it is added to the length of the pipe i.e.

$$f_1 = \frac{v}{4(L+e)}$$

for a pipe with one open end, and

$$f_1 = \frac{v}{2(L+2e)}$$

for a pipe open at both ends.

11 Current of Electricity

11.1 Charge, current and potential difference

Electric charge is the physical property of matter that causes it to experience a force when placed in an electromagnetic field. For charge to flow in a circuit, there must generally be a closed conducting path extending from the positive to negative terminals of an energy source.

The **potential difference** (p.d.) between two points in an electrical circuit is the electrical energy converted into other forms of energy per unit charge passing from one point to the other. One **volt** (1 V) is the p.d. between two points in an electrical circuit when one joule of electrical energy is converted to other forms of energy as one coulomb of charge passes from one point to the other.

Current is the rate of flow of charge through a particular cross-sectional area with respect to time.

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t} \tag{11.1}$$

The **conventional current** follows positive charges flowing from the positive to negative terminal.

The **charge** that passes through a given point is the product of the steady current flowing through that point and the time for which the current flows.

$$Q = It \tag{11.2}$$

One **coulomb** ($1\,\mathrm{C}$) is the quantity of charge that flows through a point when a steady current of one ampere flows through that point for one second.

11.2 Resistance and resistivity

The **resistance** of a device is defined as the ratio of the p.d. across it to the current flowing through it.

$$R = \frac{V}{I} \tag{11.3}$$

A resistor has a resistance of one **ohm** ($1\,\Omega$) if a current of one ampere flows through it when a p.d. of one volt is applied across it.

The power expended by a resistor is the product of the current through it and the p.d. across it.

$$P = IV = I^2R = \frac{V^2}{R} \tag{11.4}$$

Ohm's law states that the ratio of p.d. across a conductor to the current flowing through it is a constant if physical conditions like temperature remain constant.

The resistance of a metal increases with temperature. As temperature increase, vibration of the lattice ions increases; free electrons collide more frequently with the ions, experiencing more obstructions and a lower drift velocity, so resistance increases.

The resistance of a semiconductor generally decreases with temperature, as more electrons are likely to have sufficient thermal energy to escape from the ions, so there are more electrons to act as charge carriers.

A **diode** is a device that allows current to pass only in one direction. A diode with p.d. in the direction with low resistance is forward biased, and vice versa. An excessive reverse biased p.d. can cause the diode to break down (at the breakdown voltage), leading to a sudden large current i.e. a short circuit.

The resistance of a material is generally proportional to its resistivity and length, and inversely proportional to its cross-sectional area.

$$R = \frac{\rho l}{A} \tag{11.5}$$

11.3 Electromotive force and sources

The **electromotive force** of a source is the energy converted from other forms to electrical energy per unit charge delivered round a complete circuit.

$$\varepsilon = \frac{W}{Q}$$

The e.m.f. of a source is the ability of the source to convert other forms of energy to electrical energy, while p.d. across a resistor is the ability of the resistor to convert electrical energy to other forms of energy.

Sources of e.m.f. generally have an internal resistance. This resistance can simply be treated as a resistor in series with an ideal source.

For a battery with e.m.f. ε, the terminal p.d. across it is

$$V = \varepsilon - Ir \tag{11.6}$$

The **output efficiency** of a circuit is generally given by

$$\frac{R}{R+r} \tag{11.7}$$

where R is the resistance of all the devices in the circuit, and r is the internal resistance of the battery. This is at a maximum when $R = r$.

12 Gravitational Field

Newton's **law of universal gravitation** states that every particle in the universe attracts every other particle with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them i.e.

$$|\mathbf{F}| = G\frac{m_1 m_2}{r^2} \tag{12.1}$$

where G is the gravitational constant.

The law of gravitation is an inverse-square law i.e. the magnitude of force is inversely proportional to the square of something – in this case the square of the separation between the particles.

This equation holds only for point masses, but large masses at a large distance can be assumed, without losing too much accuracy, to be a point mass, so this equation is still reasonably valid.

The region in space where an object exerts a gravitational force on another object is the **gravitational field** of the first object. A gravitational field can be represented by field lines representing the resultant direction of gravitational force acting on a mass placed at any point in the field. Where the field lines are closer, the field is stronger, and vice versa.

The **gravitational field strength g** at a point in a gravitational field is the gravitational force per unit mass acting on a body placed at that point.

$$|\mathbf{g}| = G\frac{Mm}{mr^2} = G\frac{M}{r^2} \tag{12.2}$$

Near the surface of the Earth, **g** is approximately constant, as r does not vary very appreciably. However, **g** varies over different points on Earth's surface, as the Earth is not a perfect sphere – it bulges at the equator, the density of the Earth is not uniform, and the Earth is rotating about an axis through its poles – so for an object not at the poles, gravity also has to provide for centripetal acceleration.

An object is only truly weightless when it experiences no gravitational force at all. Otherwise, it experiences apparent weightlessness – it simply does not feel its weight as it experiences no normal force.

The **gravitational potential energy** U of a mass at a point in a gravitational field is the work done by an external force in bringing the mass from infinity to that point without acceleration.

$$U = -G\frac{Mm}{r} \tag{12.3}$$

At infinity, $U = 0$. The negative sign arises from the fact that the gravitational force is attractive in nature.

The **gravitational potential** Φ at a point in a gravitational field is the work done per unit mass by an external force in bringing a test mass from infinity to that point without acceleration.

$$\Phi = -G\frac{M}{r} \tag{12.4}$$

The **escape speed** is the minimum speed with which a mass should be launched from a planet's surface to escape the planet's gravitational field. It is simply the speed that gives the mass kinetic energy that makes the mass's total energy exactly equal what its total energy would be if it were stationary at infinity.

An **equipotential** line or surface is formed by all the points that have the same potential. Objects moving along an equipotential do not lose or gain any energy.

Kepler's 3rd law states that the square of an object's orbit period is proportional to the cube of its orbit radius.

A **geostationary satellite** of a planet is a satellite that orbits the planet such that it will always be above the same point on the planet. This means that the satellite must be orbiting the planet's equator in the direction of the planet's axial rotation, and the period of the orbit must be equal to the period of the rotation.

Geostationary satellites are often used for communication. They are useful because they always stay over the same point, so there is no need to adjust satellite dishes; they are also high above the Earth and so can see large areas of the Earth. However, also due to their high altitude, images taken tend to be of low spatial resolution, and because they must be over the equator, they are of limited use for latitudes more than about 70° N or S.

13 Electric Field

Coulomb's law states that the magnitude of the electric force between two point charges is directly proportional to the product of the charges and inversely proportional to the square of the distance between them i.e.

$$|\mathbf{F}| = \frac{1}{4\pi\varepsilon_0}\frac{|Q_1Q_2|}{r^2} \tag{13.1}$$

Like charges repel while unlike charges attract.

ε_0 is the permittivity of free space. The permittivity of a medium ε is usually specified relative to ε_0.

An **electric field** is a region of space in which an electric force acts on a charged particle due to the presence of some other source charge creating the field. The electric field can be represented by field lines, similar to those of gravitational fields.

The **electric field strength E** at a point in an electric field is the electrostatic force per unit positive charge exerted on a small test charge placed at the point.

$$|\mathbf{E}| = \left|\frac{\mathbf{F}}{q}\right| = \frac{1}{4\pi\varepsilon_0}\frac{|Q|}{r^2} \tag{13.2}$$

The **electric potential energy** U of a charge at a point in an electric field is the work done by an external agent in bringing the charge from infinity to that point without acceleration. For point charges,

$$U = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r} \tag{13.3}$$

A charge moving in the direction of the electric force on it loses electric potential energy, and vice versa.

The **electric potential** V at a point in an electric field is the work done per unit positive charge by an external agent in bringing a positive test charge from infinity to that point without acceleration.

$$V = \frac{U}{q} \tag{13.4}$$

For a field due to a point charge,

$$V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r} \tag{13.5}$$

Equipotentials are as in gravitational fields; they are lines or surfaces of constant potential.

The electric field between two parallel charged plates with p.d. ΔV and separation d is uniform, with electric field strength at any point

$$|\mathbf{E}| = \frac{\Delta V}{d} \tag{13.6}$$

The work done by such a field in moving a positive charge q a distance r parallel to the field is

$$W = q\Delta V\frac{r}{d}$$

Equipotentials are all lines or planes parallel to the plates, while field lines are perpendicular to the plates. Charged particles in such a field move with constant acceleration and thus in a parabolic path (neglecting gravity).

14 DC Circuits

A **direct current** is a current whose direction does not change with time.

A circuit is a network of components that forms a closed loop, allowing current to return to its source.

When components are connected in sequence i.e. with only a single current path between the components, they are connected in **series**. When components are connected such that current can flow through either one but not both, they are connected in **parallel**.

Multiple resistances connected in series and in parallel can be simplified to a single resistance with an equivalent or effective resistance.

For resistances in series,

$$R_{\text{eff}} = R_1 + R_2 + R_3 + \cdots + R_n \tag{14.1}$$

For resistances in parallel,

$$\frac{1}{R_{\text{eff}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n} \tag{14.2}$$

Kirchoff's current law states that the algebraic sum of currents into any junction is zero i.e.

$$\sum I = 0 \tag{14.3}$$

Kirchoff's voltage law states that the algebraic sum of p.d. in any loop must equal zero i.e.

$$\sum V = 0 \tag{14.4}$$

Any two points in a circuit (or between circuits) with the same potential can be considered as connected. There will be no current flowing between the two points.

14.1 Circuit components

A **voltmeter** is a circuit component that measures the p.d. across two points in a circuit. Typically, it is connected in parallel at the two points across which the p.d. is to be measured. An ideal voltmeter should have an infinite resistance so that the circuit behaves as if the voltmeter was absent. Of course, real voltmeters cannot have infinite resistance.

An **ammeter** is a circuit component that measures the current through a point in a circuit. It is connected in series at the point current is to be measured. An ideal ammeter would have zero resistance, but this is similarly impossible in reality.

A **thermistor** is a semiconductor whose resistance changes with temperature. Thermistors usually have a negative temperature coefficient such that their resistance decreases as temperature increases.

A **light-dependent resistor** (LDR) is a semiconductor whose resistance decreases as light intensity increases.

Thermistors and LDRs can be used to divide potential in a circuit, by placing another resistor in series, and then an external circuit across either resistor.

A **potentiometer** is a resistor made using a long resistance wire. Different lengths of the resistance wire will have different resistances proportional to the length, and so this can also be used to divide potential.

A potentiometer can also be used to measure a p.d. (and consequently a current) accurately (without the problems of voltmeters or ammeters) by varying the length of the potentiometer across which the p.d. is connected until there is no current between the p.d. and the potentiometer circuit (as this will be when the p.d. are equal). This length is known as the balance length, and the p.d. can be determined from there.

15 Electromagnetism

A **magnetic field** is a region of space in which a moving charge or a current-carrying conductor will experience a magnetic force when placed in the field.

The force on a moving charge in a magnetic field is

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{15.1}$$

From this, the force on a current-carrying conductor with length $\mathbf{l}$ in the direction of the conventional current I can be derived to be

$$\mathbf{F} = I\mathbf{l} \times \mathbf{B} \tag{15.2}$$

From this, it can be seen that a charge moving in a magnetic field will experience a force perpendicular to both velocity and the field, which can be predicted using Fleming's left-hand rule (or the direction of the right-handed cross product).

The **magnetic flux density** is the force acting per unit current per unit length of a conductor placed at right angles to the magnetic field i.e.

$$\mathbf{B} = \frac{\mathbf{F}}{I\mathbf{l}} \tag{15.3}$$

Magnetic flux density has units tesla (T), which is defined equivalently; one **tesla** is the magnetic flux density of a uniform magnetic field in which a straight wire carrying a current of 1 A placed perpendicular to the field experiences a force per unit length of $1\,\mathrm{N\,m^{-1}}$ in a direction at right angles to both the field and the current.

The force on a current-carrying conductor can be used to measure the magnetic flux density, through a current balance. A wire frame balanced on a pivot with a current through it is placed such that one side is in the magnetic field and experiences a magnetic force; a weight is then placed on the opposite side to balance the frame. With needed lengths, the current, and the mass of the weight, flux density can be found.

A magnetic field will never do any work on a moving charge as the force on the charge is always perpendicular to its movement. Charges moving in a magnetic field will undergo circular motion. If there is an electric field parallel to the magnetic field, charges will move in a spiral.

If an electric field and a magnetic field perpendicular to the electric field are combined, only charges moving with a particular velocity through the fields will move through the field undeflected. This can be used as a "velocity selector" of sorts.

15.1 Magnetic fields caused by currents

The magnetic field caused by a long straight wire is circular with its direction predicted by the right-hand grip rule (fingers representing the field). The field lines are concentric circles centred at the wire.

The magnetic field caused by a single circular loop is similar to that of an infinitesimally short bar magnet.

A **solenoid** is a long cylindrical coil of wire. The magnetic field due to a solenoid is similar to that of a bar magnet, and is also predicted by the right-hand grip rule (fingers representing the current).

When a ferrous core is placed within a current-carrying solenoid, it becomes temporarily magnetised as the magnetic field of the solenoid concentrates within the iron bar. The solenoid and the iron bar together produce a very strong magnetic field, stronger than produced by the solenoid alone, thus forming an electromagnet.

When two parallel conductors carry a current in the same direction, they will experience forces toward the other conductor. When two parallel conductors carry a current in the opposite direction, they will experience forces away from the other conductor. This can be predicted from first principles, using the right-hand grip rule. The force per unit length on each wire

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi d} \tag{15.4}$$

The **ampere** is defined based on the above result – one ampere is the constant current which, if maintained in two straight parallel conductors of infinite length placed 1 m in vacuum, would produce a force per unit length of $2 \times 10^{-7}\,\mathrm{N\,m^{-1}}$ between the conductors.

16 Electromagnetic Induction

Magnetic flux ϕ through a plane surface is defined as the product of the magnetic flux density normal to the surface $\mathbf{B}$ and the area $\mathbf{A}$ of the surface i.e.

$$\phi = \mathbf{B} \cdot \mathbf{A} \tag{16.1}$$

ϕ has SI units Wb. One **weber** is the magnetic flux through a surface if a magnetic field of flux density 1 T exists perpendicularly to an area of $1\,\mathrm{m}^2$.

Magnetic flux linkage Φ is the product of the number of turns N of the coil and the magnetic flux linking each turn i.e.

$$\Phi = N\phi \tag{16.2}$$

Faraday's law of induction states that the induced e.m.f. ε is directly proportional to the rate of change of magnetic flux linkage i.e.

$$\varepsilon = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{16.3}$$

Lenz's law states that the induced e.m.f. will be directed such that the current which it causes to flow opposes the change producing it. Lenz's law is a manifestation of energy conservation.

17 Alternating Currents

An alternating current is one in which the direction of current flow alternates between forward and backward with time. In the context of alternating currents, **(a) period** is the time taken to complete one complete alternation of current; **(b) frequency** is the number of complete alternations per unit time; **(c) peak value** is the maximum value; **(d) peak to peak value** is the difference between the positive and negative peak values; and **(e) root-mean-square** (r.m.s.) value is equivalent to the steady direct current that converts electrical energy to other forms of energy at the same average rate as the alternating current in a given resistance.

The r.m.s. value is the square root of the mean of the squared value. Mathematically and for current,

$$I_{rms} = \sqrt{\langle I^2 \rangle} = \frac{1}{T}\int_0^T I^2 \mathrm{d}t \tag{17.1}$$

For sinusoidal A.C.,

$$I_{rms} = \frac{I_0}{\sqrt{2}} \tag{17.2}$$

$$V_{rms} = \frac{V_0}{\sqrt{2}} \tag{17.3}$$

Since power $P = IV$,

$$\begin{aligned} P_{rms} = I_{rms}V_{rms} &= \frac{I_0}{\sqrt{2}}\frac{V_0}{\sqrt{2}} \\ &= \frac{1}{2}I_0V_0 = \frac{1}{2}P_0 \end{aligned} \tag{17.4}$$

Alternating currents can be rectified into a direct current using diodes.

17.1 Transformers

A **transformer** is a device which allows an alternating voltage to be increased or decreased, keeping the frequency the same. It consists of **(a)** a laminated soft iron core to confine the magnetic flux and ensure maximum magnetic flux linkage; **(b)** a primary winding, connected to the input A.C. supply; and **(c)** a secondary winding, connected to the output.

An ideal transformer has no resistance, loses no magnetic flux, and so is 100 % efficient. For an ideal transformer,

$$\frac{N_s}{N_p} = \frac{V_s}{V_p} = \frac{I_p}{I_s} \tag{17.5}$$

Energy losses in transformers can occur due to resistance of the windings around the iron core, flux leakage, eddy currents in the iron core, and hysteresis, which is heat generated due to the repeated reversal of the magnetisation of the core.

18 Quantum Physics

18.1 Wave-particle duality

Classical particles exhibit wavelike properties when conditions are appropriate, while classical waves i.e. light can exhibit particle-like properties. This is wave-particle duality. For any particle, the de Broglie wavelength

$$\lambda = \frac{h}{p} \tag{18.1}$$

Wave properties start to show at lengths of similar order of magnitude as the wavelength of a particle.

Phenomena like the photoelectric effect show that light has a particulate nature, while phenomena like interference and diffraction show that light has a wave nature.

A **photon** is a discrete quantum of electromagnetic energy. Photons have energy dependent on their frequency and momentum

$$E = hf = \frac{hc}{\lambda} \tag{18.2}$$

$$|\mathbf{p}| = \frac{E}{c} = \frac{h}{\lambda} \tag{18.3}$$

18.1.1 Electron diffraction

When electrons are fired at appropriate velocities at a metal foil, they diffract to form a pattern of rings which are similar to the pattern obtained when X-rays are fired at similar foil.

In this case, the metal crystals in the foil act as a diffraction grating for the electrons to diffract. A normal grating will not work as the slit is too large – as in superposition, the slit must be about the same order as the wavelength of the particles.

18.2 Photoelectric effect

The **photoelectric effect** is the phenomenon where electrons are emitted from a clean metal surface when electromagnetic radiation of sufficiently high frequency is incident on the surface. The emission of photons is instantaneous once suitable light strikes the surface. Electrons emitted through the photoelectric effect are **photoelectrons**, and the current created by the movement of electrons is the **photocurrent**.

The effect occurs because electrons absorb photons and convert their energy into kinetic energy, with which the electrons escape. The absorption of a photon in the photoelectric effect is an all-or-nothing process: either all the energy of a photon is absorbed by one electron, or none.

The **work function** of a material is the minimum energy necessary to remove a free electron from the surface of the material. This energy is used in overcoming attractive forces between an electron and the material.

For photoelectrons to be emitted, incident photons must have energy of at least the work function of the metal i.e. the photons must have frequency higher than a threshold frequency, or wavelength shorter than a threshold wavelength.

Since an electron absorbs all the energy of a photon, and the only energy loss is the work function of the metal, we have

$$hf = \Phi + T_{\text{max}} \tag{18.4}$$

i.e. the maximum kinetic energy of the photoelectrons depends only on the incident frequency and the work function.

If photoelectrons are decelerated through a potential difference of V, then the minimum potential to stop all photoelectrons is when

$$T_{\text{max}} = eV$$

and substituting into (18.4) we get

$$eV_s = hf - \Phi \tag{18.5}$$

where V_s is the stopping potential.

The intensity of a beam of photons is

$$I = \frac{E_{\text{total}}}{tA} = \frac{N}{tA} hf = nhf \qquad n = \frac{N}{tA}$$

thus an increase in intensity, holding f constant, means an increase in n i.e. there are more incident photons per unit area per unit time, and correspondingly there will be more photoelectrons and a higher photocurrent.

18.3 Atomic structure

Isolated atoms have discrete electron energy levels i.e. electrons in an atom can only possess certain specific energies. Similar to the potential of attractive fields, the energies of energy levels are negative, and infinity has zero energy.

Energy levels can be mostly taken to correspond to quantum electron shells, where $n = 1$ is the innermost quantum shell and thus the lowest energy level, and so on for $n \in \mathbb{Z}^+$.

The **ground state** of an atom is when all its electrons are at their lowest possible energy levels. The ionisation energy of an atom is the energy needed to remove an electron from the outermost energy level containing an electron to infinity (to form an ion).

Electrons can be excited from one energy level to a higher one by the absorption of a photon with energy equal to the difference in energy between the two energy levels, or through energy imparted by collision with a high-energy particle, which can have any energy as long as it is sufficient. (The particle can retain some energy and fly off.)

Electrons can de-excite from one energy level to a lower one by the emission of a photon with energy equal to the difference in energy between the two energy levels.

18.4 Line spectra

Discrete electron energy levels in isolated atoms means that an electron in an atom can only have specific energies. When an electron falls from one energy level to another, energy is released in the form of a photon with energy equal to the difference in energy between the two levels. Since the energy levels are discrete, photons of only certain frequencies can be released, hence forming a line spectrum.

For an **emission spectrum**, atoms are first excited using a high voltage or by heating. When the atoms go back to the ground state, either directly or via intermediate energy levels, photons of only certain frequencies are emitted due to the discrete energy levels, so only certain frequencies of light are observed, forming an emission spectrum i.e. discrete bright coloured lines on a dark background.

For an **absorption spectrum**, white light is used to excite atoms. Incident photons with energies exactly equal to the difference between two of an atom's energy levels are absorbed. Since energy levels are discrete, only photons of certain frequencies are absorbed. When the atoms go back to the ground state, photons of the same frequencies are emitted but in all directions. As these frequencies of light now have much lower intensity, they account for the dark lines in the absorption spectrum i.e. discrete dark lines on a continuous spectrum.

18.5 X-ray production

X-rays can be produced when a beam of high-energy electrons, created through thermionic emission by heating a metal cathode and then accelerated through a potential difference, strikes a metal anode. This produces an X-ray spectrum consisting of the continuous spectrum and the characteristic spectrum.

The **continuous spectrum** is produced when emitted electrons accelerate suddenly due to attraction from metal nuclei in the anode, emitting photons with energy equal to the change in kinetic energy, while some electrons will come to a stop and so emit photons with energy equal to their kinetic energy. Radiation produced in this way is known as bremsstrahlung i.e. braking radiation.

As electrons approach the target at different positions and angles, there is a distribution of photon energies and thus wavelengths, forming a continuous spectrum. The photon of highest energy is formed by electrons that come to a complete stop, and so we have

$$eV = hf_{\max} \tag{18.6}$$

This maximum energy is independent of the target material.

The **characteristic spectrum** is formed when emitted electrons collide into electrons bound to target atoms and cause them to be ejected, leaving vacancies that are then filled by electrons in higher energy levels, emitting a photon due to the transition, with energy equal to the difference in energy between the energy levels, creating sharp peaks in the X-ray spectrum.

When the vacancy is formed in the K ($n = 1$) shell and is filled by an electron from the L ($n = 2$) shell, the photon is termed a K_α X-ray. If the same vacancy is filled by an electron from the M shell, the photon is termed a K_β X-ray.

Generally, K_α X-rays have higher intensity than K_β X-rays as electrons in the L shell are closer to the K shell, so there is higher probability of a K shell vacancy being filled by an L electron versus an M electron. Characteristic X-rays have shorter wavelengths for targets of higher proton number, as the energy differences between shells increase.

18.6 Heisenberg's uncertainty principle

Quantum mechanics tells us that there is a limit to how precisely we can know a particle's momentum and position, given by Heisenberg's uncertainty principle. This principle also relates energy and time e.g. in the case of energy emitted by de-excitation of atoms over a specified time interval.

$$\Delta x \Delta p \geq \frac{\hbar}{2} = \frac{h}{4\pi} \tag{18.7}$$

$$\Delta E \Delta t \geq \frac{\hbar}{2} = \frac{h}{4\pi} \tag{18.8}$$

18.7 Wavefunctions

The Schrödinger's equation describes quantum systems much like Newton's laws describe classical systems. Where Newton's laws deals with position and time, Schrödinger's equation deals with wavefunctions Ψ.

The wavefunction contains all the information about a quantum system at all times. The wavefunction gives us the probability density function weighing all measurements made of the system, and from this we can derive the probability of finding a particle between positions a and b:

$$P(a \leq \mathbf{x} \leq b) = \int_a^b |\Psi|^2 \, \mathrm{d}\mathbf{x} \tag{18.9}$$

18.8 Quantum tunneling

Suppose an electron is travelling in a straight line towards a region of negative potential. The electron must do work to pass this region, and so this region is a potential barrier to the electron.

Clasically, if the electron does not have enough energy to overcome this barrier, it will never do so. Quantum mechanically, there is a probability for the electron to tunnel through the barrier even when its energy is lower than that of the potential barrier, akin to when some light passes through an interface between two materials while some light is reflected back into the material in which the incident beam is in. This is **quantum tunneling**.

When this happens, the matter wave before the barrier becomes a standing wave arising from the interference between the incident and reflected matter wave, while the wavefunction after the barrier is a wave of lower amplitude but equal energy (frequency), indicating a small probability of the electron being transmitted. Within the barrier, the wavefunction undergoes exponential decay.

The probability that an incident electron tunnels the barrier through is known as the **transmission coefficient** T, while the probability that it reflects is known as the **reflection coefficient** R. Since the electron must either be reflected or transmitted,

$$R + T = 1$$

T in this syllabus is given as

$$T \propto \exp(-2kd) \qquad k = \sqrt{\frac{8\pi^2 m(U - E)}{h^2}} \tag{18.10}$$

19 Lasers and Semiconductors

19.1 Lasers

The word '**laser**' is an acronym meaning light amplification by stimulated emission of radiation. Light emitted from a laser is monochromatic, coherent, unidirectional and focused.

19.1.1 Principles of the laser

Lasers work based on (stimulated) absorption, spontaneous emission, stimulated emission and population inversion.

Absorption occurs when an atom is excited from a lower energy level to a higher energy level due to an electron absorbing an external photon with energy equal to the energy difference between the two energy levels.

Spontaneous emission occurs when an excited atom transits on its own accord to a lower energy level, releasing a photon with energy equal to the energy difference between the initial and final energy levels.

Stimulated emission occurs when an excited atom is induced to transit to a lower energy level through interaction with an incident photon of energy equal to the energy difference between the initial and final energy levels, releasing a photon that has the same energy, phase, polarisation and direction of travel as the incident photon.

Population inversion occurs when the population of atoms at a higher energy level exceeds the population of atoms at a lower energy level. This can be achieved through optical pumping or pumping by electrical discharge.

19.1.2 Laser mechanism

The **laser mechanism** starts with an external energy source, which is used to excite atoms from ground state to an excited state. These atoms de-excite to a metastable state through spontaneous emission, where they stay for a long period of time, such that more atoms are in the higher lasing level than the lower lasing level i.e. population inversion.

When an atom in the metastable state falls back to a lower energy level, a photon is emitted. This photon interacts with another atom in the metastable state, stimulating the atom to de-excite to a lower energy state, emitting another photon in the process. This photon emitted by stimulated emission has the same phase, energy, frequency, polarization and direction of travel as the incoming photon. The incoming photon is not absorbed in the process. These photons then go on to cause other excited atoms in the metastable state to de-excite resulting in the stimulated emission of more photon.

This chain reaction will continue and light of high intensity is produced. Mirrors at both ends reflect the photons through the lasing medium, causing more stimulated emission of photons in the axis parallel to the laser tube. The partially reflective mirror at one end allows a small fraction of laser light to escape to form a useful laser beam that is coherent, collimated and monochromatic.

19.2 Semiconductors

19.2.1 Energy bands

Isolated atoms have discrete electron energy levels. However, when two atoms are brought close, the electric interaction between them causes these two energy levels to split into two levels with energies similar to the original.

When many atoms come together, like in a solid, the original discrete energy level splits into an exceedingly large number of energy levels with similar energies, which coalesce to form a continuous **energy band**.

Similar to energy levels in an isolated atom, most energy bands are either filled or empty. Like the valence shell, the **valence band** is the outermost band containing electrons. The **conduction band** refers to the next higher energy band after the valence band.

19.2.2 Band theory: electrical conduction

The energy gap between the valence and conduction band is termed the **band gap**. The width of the band gap determines whether a solid is a conductor, semiconductor or insulator.

In metals, the valence band either is incompletely filled or overlaps with the conduction band. Valence electrons can thus easily move to higher unfilled energy levels in the valence or conduction band using little to no energy, and so are free and able to move when there is an external electric field to conduct electricity. Metals are thus good conductors of electricity.

In insulators, the valence band is completely filled while the conduction band is completely vacant; there is also a large band gap. At room temperature a negligible number of atoms have sufficient energy to cross over to the conduction band. Most electrons cannot move to conduct electricity and the material is a insulator.

However, at higher temperatures or in strong electric fields, enough electrons can move to the conduction band and the material starts to conduct. When this happens, the material has 'broken down'.

A material that acts as a semiconductor when pure is an **intrinsic semiconductor**. They have conductivities and band gaps (on the order of 1 eV) between those of conductors and insulators.

At 0 K, there are no free electrons to conduct as all valence electrons are bound in the completely filled valence band and there is no energy to excite them across the energy gap, and so semiconductors are poor conductors at low temperatures.

At higher temperatures, thermal excitation of electrons across the band gap becomes more possible, and so conductivity increases.

When electrons move from the valence band to the conduction band, they become free to move when there is an external electric field in order to conduct a charge. The **hole** they leave behind in the valence band acts as a charge carrier as a nearby free electron can move into the hole, creating a new hole at the original site of the electron. The hole thus acts as a positively charged particle moving opposite to the electrons.

Both electrons and holes in the conduction and valence bands respectively are responsible for electrical conduction in semiconductors.

19.2.3 Doping

The electrical conductivity of an intrinsic semiconductor can be increased by adding impurities called **dopants**. Doped semiconductors are termed **extrinsic semiconductors**.

There are two types of dopants. **N-type dopants** are pentavalent elements from group V like phosphorus and arsenic; they are electron donors. **P-type dopants** are trivalent elements from group III like boron; they are electron acceptors.

When phosphorus is used as a dopant in silicon, its 'extra' electron (compared to silicon) is weakly bonded and can break free and move into the conduction band, helping to conduct. This type of doped semiconductor has mostly electrons as charge carriers, so it is called an n-type semiconductor.

In band theory, the addition of a donor creates **donor levels** just below the conduction band, which contains the extra electrons from donor atoms. At room temperature, lattice vibrations easily provide the small amount of energy required for donated electrons to cross to the conduction band where the electrons can conduct in the presence of an electric field.

When boron is used as a dopant in silicon, its missing electron forms a hole that can be filled by an electron from a neighbouring atom; when this happens, it creates a hole at the position the electron originated. This movement of electrons conducts electricity; equivalently, this can be seen as holes moving in the opposite direction. Since the charge carriers are mostly holes, this kind of doped semiconductor is a p-type semiconductor.

In band theory, the addition of an acceptor creates **acceptor levels** just above the valence band, representing the holes created by the acceptor atoms. At room temperature, acceptor levels are occupied by electrons thermally excited from the valence band, which leave holes in the valence band that can move to conduct electricity.

The addition of dopants does not disrupt the regular lattice of the silicon atoms, nor does it cause the material as a whole to have an overall charge.

19.3 p-n junction

A **p-n junction** is formed when a p-type conductor is joined to an n-type. The difference in concentration of electrons between the n-type and p-type causes electrons to diffuse across the junction from the n-type filling the holes in the p-type. This depletes the electrons and holes in the n-type and p-type respectively, creating a **depletion region** almost devoid of charge carriers.

When electrons diffuse away from the n-type, they leave behind immobile cations, while the filling of holes in the p-type creates anions. This results in an internal electric field being created across the depletion region.

As the diffusion continues, the depletion region widens and the charge difference across the junction increases, and eventually the internal electric field becomes so strong that further diffusion across the junction is prevented, an an equilibrium is reached.

When the p-type is connected to the positive terminal and the n-type to the negative terminal of a voltage source, the p-n junction is under forward bias. The internal potential difference decreases as its polarity is opposite to that of the external voltage source, and so the depletion region becomes narrower and stops inhibiting the flow of electrons from n-type to p-type. Electrons can now flow to conduct the current and current increases exponentially with increasing forward voltage.

When the p-type is connected to the negative terminal and the n-type is connected to the positive terminal of a voltage source, the p-n junction is under reverse bias. The internal potential difference increases with increasing reverse bias, widening the depletion region and further inhibiting the flow of current across the junction. Only a small reverse current can flow through, carried by the small number of n-type electrons in the conduction band and p-type holes in the valence band.

The p-n junction thus conducts current in one direction and resists current in the other; this can be used to rectify alternating currents. When the AC is flowing in the forward bias direction of the diode, the diode has low resistance and allows the current to flow. When the AC is flowing in the reverse bias direction of the diode, the diode has very high resistance, only allowing negligible current to flow. Effectively, the current is forced to flow in only one direction.

20 Nuclear Physics

20.1 α particle experiment

In Rutherford's α particle scattering experiment, α particles were fired at a thin gold foil.

Most α particles made it through the foil undeflected or with a small deflection. This showed that the probability of an α getting close to the positive charge in the nucleus is small, which indicated that the nucleus occupies only a small amount of space.

Some α were deflected through large angles, indicating that the atom must have a positively-charged nucleus of very small dimensions in order to provide the large force required for such a deflection.

Few (but not zero) α were reflected backwards. This further proved that the nucleus was small and very massive.

20.2 Nuclear structure

The nucleus contains positive protons and neutral neutrons, which are collectively known as neutrons. Compared to the atom which has diameter on the order of 1×10^{-10} m, the nucleus has diameter on the order of 1×10^{-15} m. Most of the mass of an atom is in the nucleus.

Atoms consist of a positive and small nucleus orbited by negative electrons in defined orbitals and a lot of empty space.

The **proton number** or **atomic number** Z is the number of protons in a nucleus. The **nucleon number** or **mass number** $A = Z + N$ is the number of nucleons in a nucleus. N is the number of neutrons. A **nuclide** is a particular species with a unique proton number and mass number represented by ${}^{A}_{Z}\mathrm{X}$ e.g. carbon-12: ${}^{12}_{6}\mathrm{C}$.

Isotopes are atoms that have the same number of protons but different number of neutrons. The abundance of isotopes is generally constant throughout the natural world.

20.3 Atomic mass unit

Atomic masses are usually expressed in terms of the **atomic mass unit** (u). 1 u is one-twelfth the mass of the carbon-12 atom.

Masses can also be expressed as relative atomic masses. The relative atomic mass of an atom is the ratio of the mass of the atom to the unified atomic mass unit; it is numerically equal to the atomic mass, but is unitless.

20.4 Mass-energy equivalence

As a result of special relativity, mass and energy are equivalent, and their conservation laws can be unified into the law of conservation of mass-energy. Mass and energy are related by

$$E = mc^2 \tag{20.1}$$

In nuclear physics (and quantum physics) the **electron volt** (eV) is commonly used. 1 eV is the energy gained by a charge equal to that on an electron in moving through a potential difference of 1 V.

20.5 Binding energy

When separate nucleons come together to form a nucleus, the resulting nucleus has less mass than the separate nucleons as some of the mass has been converted into potential energy holding the nucleons together. This energy is known as the binding energy.

Formally, the **binding energy** of a nucleus is the minimum energy required to completely separate the nucleus into its constituent nucleons and protons.

The difference in mass caused by binding energy is known as the mass defect. Formally, the **mass defect** of a nucleus is the difference between the mass of the separated nucleons and the combined mass of the nucleus.

The mass defect can easily be calculated as in the definition, and the binding energy from the mass defect.

$$\begin{aligned} \Delta M &= Zm_p + Nm_n - M_{\text{nucleus}} \\ &= Zm_p + Nm_n + Zm_e - M_{\text{atom}} \end{aligned} \tag{20.2}$$

$$E_B = \Delta Mc^2 \tag{20.3}$$

The binding energy per nucleon is an indicator of the stability of the nucleus. Naturally, the more binding energy per nucleon, the more stable the nucleus.

In general, the binding energy per nucleon increases sharply from the lighter elements and peaks at iron before falling gradually. This implies that iron is the most stable element – and it is.

20.6 Nuclear processes

In nuclear reactions, nucleons are rearranged similar to how atoms are rearranged in chemical reactions. In all nuclear processes, nucleon number, proton number i.e. charge, momentum and mass-energy must be conserved.

Nuclear reactions can be represented by equations just like chemical reactions e.g. ${}^{1}_{0}\mathrm{n} + {}^{14}_{7}\mathrm{N} \longrightarrow {}^{14}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{H}$.

The previous reaction can also be represented as $^{14}_{7}\text{N}$ (n, p) $^{14}_{6}\text{C}$.

To find the energy released in a nuclear process, just start with conservation of energy, which leads you to

$$\begin{aligned} E_{\text{reactants}} &= E_{\text{products}} + E_{\text{released}} \\ \Rightarrow E_{\text{released}} &= E_{\text{reactants}} - E_{\text{products}} \\ &= E_{B\,(\text{products})} - E_{B\,(\text{reactants})} \end{aligned} \tag{20.4}$$

If the products have more energy than the reactants, then energy must be supplied in the form of kinetic energy for the reaction to occur. Otherwise, the reaction can (theoretically) occur at rest as long as all conservations are satisfied.

Energy released in nuclear reactions is carried away either as kinetic energy or by a photon.

20.6.1 Nuclear fusion and fission

Nuclear fission is the disintegration of a heavy nucleus into two lighter nuclei of approximately equal masses. Typically, energy is released as the binding energy of the heavy nucleus is less than that of the fission products.

Nuclear fusion is the combination of two light nuclei to produce a heavier nucleus. Typically, energy is released as the binding energy of the reactants is less than that of the fusion product.

Atoms of elements before iron generally undergo nuclear fusion as they have less binding energy per nucleon than iron. Atoms of elements after iron generally undergo nuclear fission as they have less binding energy per nucleon than iron.

20.7 Nuclear decay

Radioactive decay is the spontaneous disintegration of the nucleus of an atom resulting in the emission of particles or radiation. Radioactive decay is a spontaneous and random process.

A **spontaneous process** cannot be sped up or slowed down by physical means and is independent of any chemical condition and the decay of other atoms.

Being a **random process** means that it is impossible to predict which nucleus and when any particular nucleus will disintegrate.

20.7.1 Types of radiation

α particles are simply high-energy helium-4 nuclei – $^{4}_{2}\text{He}$. When emitted from α decay, they usually possess energies in MeV range. They have high ionising power and a range of about 3 cm to 4 cm, and are easily stopped by a sheet of paper.

β particles are high-energy electrons emitted through β(-minus) decay, when a neutron decays into a proton and an electron. They have speeds up to $0.5c$. Their ionising power is about a tenth of that of α particles, but they have a range about 10 times that of alpha particles. They can only be stopped by a few mm of aluminium.

γ rays are simply high-energy photons, with frequencies higher than X-ray frequency. γ rays have ionising power about one-ten thousandth that of α particles, but they have the longest range, only being stopped by a few cm of lead. Gamma decay represents the emission of energy when an excited nucleus returns to its ground state.

20.7.2 Decay law

The rate of radioactive decay of a sample is proportional to the number of radioactive nuclei present. Mathematically,

$$-\frac{\mathrm{d}N}{\mathrm{d}t} \propto N \Rightarrow \frac{\mathrm{d}N}{\mathrm{d}t} = -\lambda N$$

Solving, this gives us

$$N = N_0 e^{-\lambda t} \tag{20.5}$$

The **decay constant** λ of a nucleus is the probability of decay per unit time of the nucleus.

The **activity** of a sample is the average number of atoms disintegrating per unit time.

$$\begin{aligned} A &= -\frac{\mathrm{d}N}{\mathrm{d}t} = \lambda N \\ &= \lambda N_0 e^{-\lambda t} = A_0 e^{-\lambda t} \end{aligned} \tag{20.6}$$

The **half-life** of a radioactive substance is the time taken for half the original number of radioactive nuclei to decay. From (20.5), we can derive

$$t_{1/2} = \frac{\ln 2}{\lambda} \tag{20.7}$$

$$\frac{N}{N_0} = \frac{A}{A_0} = \left(\frac{1}{2}\right)^{t/t_{1/2}} = \left(\frac{1}{2}\right)^{n} \tag{20.8}$$

where n is the number of half-lives.

20.7.3 Background radiation

When dealing with experimental measurements of a radioactive source, we must take into account background radiation.

Background radiation comes from various sources, including **(a)** the air; **(b)** building materials; **(c)** the soil; **(d)** water; **(e)** human bodies; and **(f)** medical sources.

20.7.4 Radiation's effects on biological organisms

Radiation hazards to biological organisms arise from exposure, ingestion and inhalation. Doses of radiation are measured in sieverts (Sv).

Severe doses of radiation can lead to immediate effects like radiation burns, and delayed effects like cancer, eye cataracts and hereditary defects due to genetic damage.

Generally, α radiation causes little damage as it can barely penetrate the skin; β radiation causes some damage but can be stopped by surface tissues; and γ radiation can penetrate deeply, causing damage to organs deep in the body.