Physics

https://asdia.dev/notes/physics.pdf

Notes taken by
Eytan Chong

2024–2025

Acknowledgements. A significant portion of this work is derived from notes provided by the Physics Department at Dunman High School, and *A-Level Study Guide – Physics – Edition H2.2* by CS Toh.

Last updated September 9, 2025.

Contents

Part I

Measurement

1 Measurement

1.1 Physical Quantities and SI Units

Definition 1.1.1. A *physical quantity* is a scientifically measurable quantity, and consists of a numerical value and a unit.

Physical quantities can be classified as *base* or *derived* quantities, and likewise for units.

Definition 1.1.2. A *base quantity* is one that is not dependent on other quantities. The units of base quantities are called *base units*.

The seven SI base quantities (and their units) are:

Base Quantity	**Base Unit**	**Symbol**
Mass	Kilogram	kg
Length	Metre	m
Time	Second	s
Temperature	Kelvin	K
Amount of Substance	Mole	mol
Electric Current	Ampere	A
Luminous Intensity	Candela	cd

Figure 1.1: The seven SI base quantities and units.

Definition 1.1.3. A *derived quantity* is one that is expressed as a product/quotient of base quantities. A derived quantity has a *derived unit* that is expressed as a product/quotient of base units.

Definition 1.1.4. An equation is *homogeneous* if all terms have the same units.

A physically correct equation must be homogeneous.

1.2 Scalar and Vectors

Definition 1.2.1. *Scalars* are quantities that have magnitude only.

Definition 1.2.2. *Vectors* are quantities that have both magnitude and direction.

Geometrically, a vector can be represented with an arrow whose direction represents its relative direction and whose length represents its relative magnitude.

For vector algebra, see here.

1.3 Errors and Uncertainties

Definition 1.3.1. *Uncertainty* is the range of values on both sides of a measurement in which the actual value of the measurement is expected to lie.

Uncertainties in measured quantities arise from

- limitations of the observer;
- limitations of the measuring instrument used;
- limitations of the method used.

Definition 1.3.2. *Error* is the difference between the measured value and the true value.

There are two types of errors: systematic and random.

Definition 1.3.3. *Systematic errors* are constant deviations of the readings in one direction from the true value.

Systematic errors can be eliminated by careful experimental design and good experimental techniques.

Definition 1.3.4. *Random errors* refer to the scatter of readings about a mean value (usually the sum of the true value and all systematic errors).

Random errors cannot be eliminated, but their effects can be reduced by taking the average of repeated readings.

1.3.1 Precision and Accuracy

Definition 1.3.5. *Precision* is the degree of agreement among repeated measurements of the same quantity.

Good precision is associated with small random errors.

Definition 1.3.6. *Accuracy* is the degree of agreement between measurements and the true value.

Good accuracy is associated with small systematic errors.

1.3.2 Derived Uncertainties

Definition 1.3.7. The uncertainty Δx in the value $x \pm \Delta x$ of a quantity is also called its *absolute uncertainty*. The corresponding *relative uncertainty* is given by $\Delta x/x$.

- If $Z = mA + nB$, then
$$\Delta Z = |m|\,\Delta A + |n|\,\Delta B.$$
- If $Z = kA^m B^n$, then
$$\frac{\Delta Z}{Z} = |m|\,\frac{\Delta A}{A} + |n|\,\frac{\Delta B}{B}.$$
- Else, we compute ΔZ using
$$\Delta Z = \frac{Z_{\text{max}} - Z_{\text{min}}}{2}.$$

Part II

Newtonian Mechanics

2 Kinematics

2.1 Rectilinear Motion

2.1.1 Time Derivatives of Position

Definition 2.1.1. The *distance* (d) travelled by a body is the length of the path it has taken.

Distance is a scalar quantity. Its SI unit is metre (m).

Definition 2.1.2. *Displacement* (s, x) is the straight-line distance and the direction from a fixed point.

Displacement is a vector quantity. Its SI unit is metre (m).

Definition 2.1.3. *Speed* (v) is the distance travelled per unit time.

Speed is a scalar quantity. Its SI unit is metre per second (m s^{-1}).

Definition 2.1.4. *Velocity* (v) is the rate of change of displacement.

$$v = \frac{\mathrm{d}x}{\mathrm{d}t}.$$

Velocity is a vector quantity. Its SI unit is metre per second (m s^{-1}).

Definition 2.1.5. *Acceleration* (a) is the rate of change of velocity.

$$a = \frac{\mathrm{d}v}{\mathrm{d}t}.$$

Acceleration is a vector quantity. Its SI unit is metre per second square (m s^{-2}).

2.1.2 Graphical Methods

Given a displacement-time (s-t) graph, the velocity v is given by the gradient of the graph.

Given a velocity-time (v-t) graph,

- the change in displacement Δs in any time interval Δt is given by the corresponding signed area under the graph; and
- the acceleration is given by the gradient of the graph.

2.1.3 Equations of Motion

Theorem 2.1.6 (Equations of Motion)**.** Consider a body moving with uniform acceleration a in a straight line. If its initial velocity is u, and its final velocity at time t is v, then

1. $v = u + at$,
2. $s = \dfrac{u+v}{2}t$,
3. $s = ut + \dfrac{1}{2}at^2$,
4. $s = vt - \dfrac{1}{2}at^2$,
5. $v^2 = u^2 + 2as$.

2.1.4 Free-Fall

Definition 2.1.7. *Free-falling* refers to motion under the influence of a uniform gravitational field without air resistance.

A free-falling body undergoes uniformly accelerated motion in the vertical direction. Using the equations of motion, we see that (taking the upward direction as positive):

$$s = ut - \frac{1}{2}gt^2 \quad \text{and} \quad v = gt,$$

hence the body follows a parabolic trajectory, and the speed of the body will increase uniformly as it falls.

For a body projected upward, the motion has a *speed symmetry*, i.e. the speed v during the upward motion is equal to the speed v at the same point during the downward motion. Accordingly, there would also be a *time symmetry*, i.e. the time Δt for the body to reach maximum height from any point is equal to the time Δt for it to return to the same point.

Definition 2.1.8. *Air resistance* (R) is the resistive force acting on a body moving through air.

The direction of air resistance is always opposite that of motion. At low velocities, $R \propto v$. At high velocities, $R \propto v^2$.

If the body remains long enough in the air, eventually $R \to W$, and it will no longer accelerate. The body has reached *terminal velocity*.

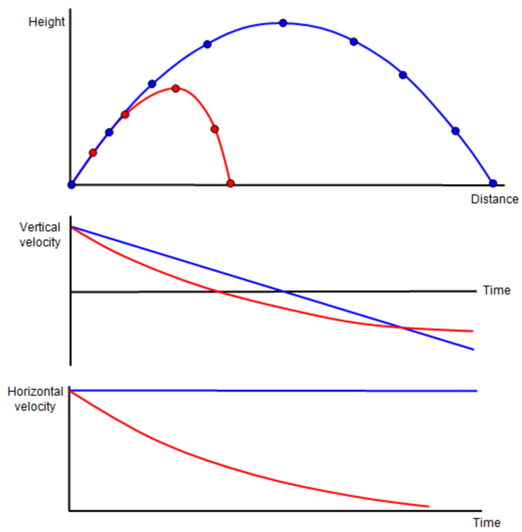

Figure 2.1: The s-t and v-t graphs for negligible and non-negligible air resistance.[1]

[1]Source: https://www.schoolphysics.co.uk/age16-19/Mechanics/Kinematics/text/Projectiles_and_air_resistance/index.html

2.2 Projectile Motion

Definition 2.2.1. *Projectile motion* refers to the motion of a body in which there is a uniform velocity in one direction and a uniform acceleration in a perpendicular direction.

Consider an idealized model, where the projectile is a particle with acceleration (due to gravity) that is constant in both magnitude and direction. Note that the plane of motion is two-dimensional since the acceleration due to the gravity is purely vertical (the projectile cannot move sideways). We represent this plane of motion with the xy-plane. The trajectory of the object depends only on the initial velocity u and the downward acceleration due to gravity g.

By treating the x- and y-coordinates separately, we can easily analyse projectile motion. With initial speed u and projection angle α, we obtain the following formulae:

- In the *horizontal direction*, there is no resultant force, so $a_x = 0$. By the equations of motion, we get

$$\begin{aligned} a_x &= 0, \\ v_x &= u_x \\ &= u\cos\alpha, \\ s_x &= u_x t \\ &= (u\cos\alpha)\,t. \end{aligned}$$

- In the *vertical direction*, there is a constant acceleration of $-g$. Taking the upwards direction as positive, we get

$$\begin{aligned} a_y &= -g, \\ v_y &= u_y - gt \\ &= u\sin\alpha - gt, \\ s_y &= u_y t - \frac{1}{2}gt^2 \\ &= (u\sin\alpha)\,t - \frac{1}{2}gt^2. \end{aligned}$$

The magnitude v and direction θ of the velocity of the particle at any instant can be calculated using

$$v = \sqrt{v_x^2 + v_y^2} \quad \text{and} \quad \theta = \arctan\frac{v_y}{v_x}.$$

Proposition 2.2.2. The time to reach the maximum height is $u\sin\alpha/g$.

Proof. At maximum height, $v_y = 0$. Thus,

$$u\sin\alpha - gt = 0 \implies t = \frac{u\sin\alpha}{g}.$$

□

Using time symmetry, we easily obtain the total flight time of the body.

Corollary 2.2.3. The total time of flight is $2u\sin\alpha/g$.

Proposition 2.2.4. The maximum height attained by the particle is $u^2 \sin^2 \alpha/2g$.

Proof. We know the body is highest when $t = u \sin \alpha/g$. Substituting this into the equation for s_y, we get

$$s_y = (u \sin \alpha) \left(\frac{u \sin \alpha}{g}\right) - \frac{1}{2} g \left(\frac{u \sin \alpha}{g}\right)^2 = \frac{u^2 \sin^2 \alpha}{2g}.$$

□

Proposition 2.2.5. The range (maximum horizontal distance covered by the trajectory) is u^2/g.

Proof. We know that it takes $t = 2u \sin \alpha/g$ for the object to land. Substituting this into the equation for s_x,

$$s_x = (u \cos \alpha) \left(\frac{2u \sin \alpha}{g}\right) = \frac{u^2 \sin 2\alpha}{g}.$$

Clearly, s_x is maximized when $\sin 2\alpha = 1$ (i.e. $\alpha = \pi/4$), whence the range is u^2/g. □

3 Dynamics

3.1 Newton's Laws of Motion

Law 3.1.1 (Newton's Laws of Motion).

1. An object at rest will remain at rest and an object in motion will remain in motion at constant velocity in the absence of an external resultant force.
2. The rate of change of momentum of a body is directly proportional to the resultant force acting on the body and occurs in the direction of the resultant force.
3. If body A exerts a force on body B, then body B exerts a force of the same type that is equal in magnitude and opposite in direction on body A.

3.1.1 Mass and Weight

Newton's first law suggests that matter has a property called inertia.

Definition 3.1.2. *Inertia* is the resistance to acceleration upon the application of a force.

Definition 3.1.3. The *mass* of a body is a measure of its inertia.

Mass is a scalar quantity. Its SI unit is kilogram (kg).

Definition 3.1.4. *Weight* is the force experienced by a mass in a gravitational field.

Weight is a vector quantity. Its SI unit is newton (N).
The weight W of a body is related to its mass m and the acceleration of free fall g by

$$W = mg.$$

3.1.2 Momentum, Force and Impulse

Definition 3.1.5. The *linear momentum* (p) of a body is the product of its mass m and its velocity v:

$$p = mv.$$

Momentum is a vector quantity. Its SI unit is kilogram-metre per second (kg m s^{-1}). The direction of momentum is the same as the direction of the velocity.
In accordance with Newton's second law,

Definition 3.1.6. *Force* (F) is the rate of change of momentum of the body acted upon.

$$F = \frac{\mathrm{d}p}{\mathrm{d}t}.$$

Force is a vector quantity. Its SI unit is newton (N).
In the case of *constant mass*, then

$$F = \frac{\mathrm{d}(mv)}{\mathrm{d}t} = m\frac{\mathrm{d}v}{\mathrm{d}t} = ma.$$

Definition 3.1.7. *Impulse* (J) is the product of the force F and the time interval Δt during which the force acts.

$$J = F\Delta t = \int F \, \mathrm{d}t.$$

Impulse is a vector quantity and has the same direction as the applied force. Its SI unit is newton-second (N s).

Theorem 3.1.8 (Impulse-Momentum Theorem)**.** Impulse is equal to the change in momentum.

Proof. By Newton's second law, we have

$$J = \int F \, \mathrm{d}t = \int \frac{\mathrm{d}p}{\mathrm{d}t} \, \mathrm{d}t = \int \mathrm{d}p = \Delta p.$$

□

3.1.3 Action-Reaction Pairs

Newton's third law implies that forces occur in action-reaction pairs.

Definition 3.1.9. An *action-reaction pair* of forces are two forces which fulfil the following conditions:

- the two forces act on two different bodies,
- the forces are equal in magnitude,
- the forces are opposite in direction, and
- the forces are of the same type.

3.2 Conservation of Momentum

Law 3.2.1 (Conservation of Momentum)**.** The total momentum of a system is constant, provided no resultant external force acts on the system.

Justification. Let body A have mass m_1 and initial velocity u_1. Let body B have mass m_2 and initial velocity u_2. Suppose bodies A and B collide. During the collision, body A exerts a force F_2 on body B, and body B exerts a force F_1 on body A. By Newton's third law, these forces are equal and opposite:

$$F_2 = -F_1.$$

Both forces act for the same time Δt, hence the impulses of the two forces are equal and opposite:

$$J_2 = F_2\Delta t = -F_1\Delta t = -J_1.$$

Since $J = \Delta p = m\Delta v$, we have

$$m_2 \left(v_2 - u_2\right) = -m_1 \left(v_1 - u_1\right).$$

Rearranging, we get

$$m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2,$$

so momentum is conserved. □

3.2.1 Elastic Collisions

Definition 3.2.2. In an *elastic collision*, total kinetic energy is conserved.

Proposition 3.2.3. In an elastic collision, the relative speed of approach of the two bodies is equal to the relative speed of separation.

Proof. Consider two bodies A and B, with mass m_1 and m_2 respectively, colliding head-on elastically with each other. The initial speeds of A and B are u_1 and u_2 respectively. After the collision, A and B move off with final speeds v_1 and v_2 respectively.

By the law of conservation of momentum,

$$m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2.$$

Rewriting, we see that

$$m_1 (u_1 - v_1) = m_2 (v_2 - u_2). \tag{3.1}$$

Since kinetic energy is conserved,

$$\frac{1}{2} m_1 u_1^2 + \frac{1}{2} m_2 u_2^2 = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2.$$

Rewriting,

$$m_1 (u_1 + v_1)(u_1 - v_1) = m_2 (v_2 + u_2)(v_2 - u_2). \tag{3.2}$$

Dividing (3.2) by (3.1), we obtain

$$u_1 + v_1 = v_2 + u_2 \implies u_1 - u_2 = -(v_1 - v_2),$$

so the relative speed of approach is equal to the relative speed of separation as desired. □

3.2.2 Inelastic Collisions

Definition 3.2.4. In an *elastic collision*, total kinetic energy is not conserved. In a *completely inelastic collision*, the colliding bodies stick together and move off as one body after the collision.

4 Forces

4.1 Types of Forces

4.1.1 Force Fields and Conservative Forces

Definition 4.1.1. A *force field* is the region of space within which a force is experienced.

Definition 4.1.2. A force is *conservative* if the work it does on an object moving between two points is independent of the path taken by the object. Else, the force is *non-conservative*.

The gravitational force is an example of a conservative force, while friction is an example of a non-conservative force.

4.1.2 Tension

Definition 4.1.3. *Tension* (T) is the force exerted by an extended body on another body to which it is attached, and acts along the direction of the extended body.

Law 4.1.4 (Hooke's Law)**.** Within its elastic limit, the magnitude of the tension in an elastic body is directly proportional to its extension.

Mathematically,

$$T = kx,$$

where k is the spring/force constant and x is the extension.

For a given elastic material, the spring constant k is directly proportional to its cross-sectional area A and inversely proportional to its natural length l_0:

$$k \propto A \quad \text{and} \quad k \propto \frac{1}{l_0}.$$

If two springs of spring constant k_1 and k_2 are connected in parallel, they can be replaced by a single spring with spring constant

$$k_{\text{eff}} = k_1 + k_2.$$

If they are connected in series, they can be replaced by a single spring of spring constant

$$\frac{1}{k_{\text{eff}}} = \frac{1}{k_1} + \frac{1}{k_2}.$$

4.1.3 Normal Contact Forces and Frictional Forces

Definition 4.1.5. A *contact force* is a force that can only act when objects are physically touching.

When two solid surfaces are in contact, it is usually possible to represent the forces between them as two components.

- *normal contact forces* that are perpendicular to the surfaces, and
- *frictional forces* that are parallel to the surfaces.

Frictional forces are always opposite in the direction to motion.

4.1.4 Viscous Forces

Definition 4.1.6. *Viscous force* is the frictional force experienced by a body when it moves in a fluid.

The origin of viscous forces is frictional drag by the fluid and collisions with the displaced molecules of the fluid.

Unlike frictional forces, viscous forces are zero when the body's velocity is zero.

As the relative velocity of the body in the fluid increases, the viscous force on the body increases.

4.1.5 Upthrust

Definition 4.1.7. *Pressure* (p) is the normal force per unit area.

$$p = \frac{F}{A}.$$

Pressure is a scalar quantity. Its SI unit is pascal (Pa) or the newton per square metre. Alternate units include the atmosphere (atm), the millimetre of mercury (mmHg) and the bar.

Definition 4.1.8. *Density* (ρ) is the mass per unit volume.

$$\rho = \frac{m}{V}.$$

Density is a scalar quantity. Its SI unit is kilogram per cubic metre (kg m^{-3}).

Proposition 4.1.9 (Hydrostatic Pressure). Consider an object at depth h in a fluid of density ρ. The pressure p exerted *by the fluid* on the object is given by

$$p = \rho g h.$$

Proof. By the definitions of pressure and density, we have

$$p = \frac{F}{A} = \frac{\text{weight of fluid column}}{A} = \frac{mg}{A} = \frac{\rho V g}{A} = \rho g h.$$

□

This implies that pressure is independent of the cross-sectional area of the body and the shape of the container holding the fluid.

If the surface of the fluid is subject to pressure p_s from another fluid, then the pressure p_s should be added:

$$p = \rho g h + p_s.$$

Definition 4.1.10. *Upthrust*, U is the vertical upward force exerted by a fluid on a body when the body is (completely or partially) submerged in the fluid.

Upthrust arises because pressure in a fluid increases with depth, hence pressure on the lower surface is greater than that on the upper surface, resulting in a net upward force.

Principle 4.1.11 (Archimedes' Principle). When a body is immersed in a fluid, it experiences an upthrust equal in magnitude to the weight of fluid displaced. Mathematically,

$$U = m_f g,$$

where m_f is the mass of the fluid.

Justification. Consider a cylinder of cross-sectional area A and length l that is submerged in a fluid of density ρ.

The net upward force U is thus given by

$$U = \Delta F = \Delta(pA) = \Delta(\rho g h A) = (\rho A \Delta h)\, g = (\rho A l)\, g = (\rho V)\, g = m_f g.$$

□

Principle 4.1.12 (Principle of Floatation)**.** When an object is floating in a fluid, the weight of the object is equal to the weight of fluid displaced by the object.

4.2 Turning Effect of Forces

The motion of a body is made up of translation and rotation. A motion is

- purely *translational* if every particle in the body has the same instantaneous velocity,
- purely *rotational* if every particle is in circular motion about the same axis of rotation.

4.2.1 Centre of Gravity

Definition 4.2.1. The *centre of gravity* (C.G.) of a body is the point at which the whole weight of the body may be considered to act.

Definition 4.2.2. The *centre of mass* of a body is the point at which the mass of the body appears to be concentrated.

In a uniform gravitational field, the C.G. coincides with the centre of mass.

4.2.2 Moment and Torque

Definition 4.2.3. The *moment* (M) of a force F about a pivot point is the product of the force and the perpendicular distance d between its line of action and the pivot point.

$$M = Fd.$$

Moment is a vector quantity. Its direction is given by the right-hand rule. Its SI unit is newton-metre (N m).

Intuitively, moment can be thought of as the ability of a force to rotate a body about a given pivot point.

Definition 4.2.4. A *couple* is a pair of forces of equal magnitude but opposite in directions.

A couple tends to produce rotations only.

Definition 4.2.5. The *torque* (τ) of a couple is the product of one of the forces F and the perpendicular distance d between the forces.

$$\tau = Fd.$$

Torque is a vector quantity. Its SI unit is newton metre (N m).

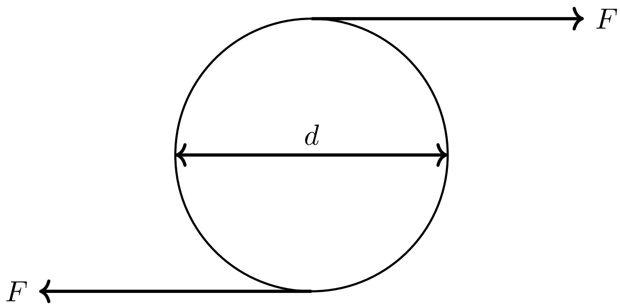

Figure 4.1: A couple. The magnitude of the torque is given by Fd.

Any system of forces acting on a body can usually be reduced to

- a single force acting through a point which only affects translational motion, and
- a torque (of a couple) which only affects its rotational motion.

On their own, the term "moment" usually refers to a turning effect of a force, while the term "torque" usually refers to a turning effect without translational effect.

4.3 Equilibrium of Forces

Definition 4.3.1.

- A body is in *translational equilibrium* if the resultant force on it is zero in all directions.
- A body is in *rotational equilibrium* if the resultant torque on it is zero about all axes of rotation.
- A body is said to be in *equilibrium* if it is in both translational and rotational equilibrium.

Principle 4.3.2 (Principle of Moments)**.** When a system is in equilibrium, the sum of clockwise moments about any axis must be equal to the sum of anti-clockwise moments about the same axis.

If three (non-parallel) forces result in equilibrium, then their lines of action share a common point.

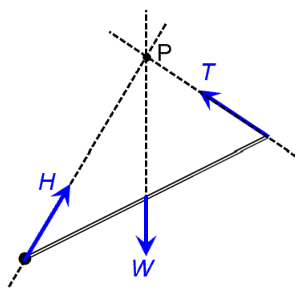

Figure 4.2: The lines of action of H, T and W intersect at a common point P.

5 Work, Energy and Power

5.1 Work and Energy

Definition 5.1.1. *Work* (W) is done when a force moves its point of application in the direction of the force. Mathematically,

$$W = Fs\cos\theta = \int F \cdot \mathrm{d}s,$$

where the displacement s is at an angle θ to the direction of the force F.

Work is a scalar quantity. Its SI unit is the joule (J).

5.1.1 Work Done by a Gas

Proposition 5.1.2. The work done W by a gas expanding its volume by ΔV against an external pressure p is given by

$$W = p\Delta V = \int p \,\mathrm{d}V.$$

Proof. We have $W = Fs = pAs = p\Delta v$. □

5.2 Energy

Definition 5.2.1. *Energy* is the ability to do work.

Energy is a scalar quantity. Its SI unit is the joule (J).

We can think of work as an energy transfer. If the work done on a system is positive, energy is transferred to the system. If the work done is negative, energy is transferred from the system.

Law 5.2.2 (Conservation of Energy)**.** Energy can neither be destroyed nor created in any process, but it can be converted from one form to another, and transferred from one body to another.

The law of conservation of energy implies that the total energy output of a machine should equal its energy input. In practice however, the *useful* energy output of a machine is usually less than its energy input. This is because some work has to be done against dissipative forces (e.g. friction).

Definition 5.2.3. The *efficiency* (η) of a machine is measured by the ratio

$$\eta = \frac{\text{useful energy output}}{\text{energy input}}.$$

5.2.1 Kinetic Energy

Definition 5.2.4. The *kinetic energy* (KE, E_k) of a body is its ability to do work due to (or as a result of) its motion.

Kinetic energy can be taken as equal to the work done to bring the body from rest to speed.

Proposition 5.2.5. The kinetic energy E_k of a body of mass m moving at speed v is given by

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

Proof. Consider a body of mass m accelerated from rest to speed v by a constant force F moving it through a displacement s.

Since F is constant, there is constant acceleration. Invoking the equations of motion, we see that

$$v^2 = 0^2 + 2as \implies s = \frac{v^2}{2a}.$$

Thus, E_k, which is equal to the work done to accelerate the body, is given by

$$E_k = Fs = (ma)\left(\frac{v^2}{2a}\right) = \frac{1}{2}mv^2.$$

□

Theorem 5.2.6 (Work-Energy Theorem)**.** The work done by a net force on a body is equal to the change in kinetic energy of the body.

$$W = \Delta E_k.$$

5.2.2 Potential Energy

Definition 5.2.7. The *potential energy* (PE, U) of a body is its ability to do work due to (or as a result of) its position and/or shape.

In a force field, the direction of the field's force is always towards lower potential energy. If a body is moved in directions perpendicular to the field, its potential energy will remain the same.

The force experienced by a stationary body at a point in a force field is numerical equal (but opposite in sign) to the potential energy gradient at that point, i.e.

$$F = -\frac{\mathrm{d}U}{\mathrm{d}x}.$$

5.2.3 Gravitational Potential Energy

Definition 5.2.8. The *gravitational potential energy* (GPE, E_p) of a body is its ability to do work due to its position.

Gravitational potential energy can be taken as equal to work done against gravity to elevate the body from some reference plane to that position.

Proposition 5.2.9. The gravitational potential energy E_p of a body of mass m at a height h (above some reference plane near the Earth's surface) is given by

$$E_p = mgh.$$

Proof. Consider a body of mass m elevated without acceleration by a constant force F through a vertical displacement h near the Earth's surface.

Since there is no acceleration,

$$F - W = 0 \implies F = mg.$$

Hence, the gravitational potential energy may be calculated as

$$E_p = Fh = mgh.$$

□

Here, h is assumed to be small compared to the Earth's radius so that the value of g may be taken as constant.

5.2.4 Elastic Potential Energy

Definition 5.2.10. The *elastic potential energy* (EPE, E_s) of an elastic body is its ability to work due to its changed shape.

Elastic potential energy may be taken as equal to the work done to change the shape of the body.

Proposition 5.2.11. The elastic potential energy E_s of an elastic body with spring constant k and extension x is given by

$$E_s = \frac{1}{2}kx^2.$$

Proof. Consider an elastic body of spring constant k extended by an external force F to an extension x. By Hooke's law, $F = kx$, so

$$E_s = \int F \,\mathrm{d}x = \int kx \,\mathrm{d}x = \frac{1}{2}kx^2.$$

□

E_s is also given by the signed area under the F-x graph.

5.2.5 Mechanical Energy

Definition 5.2.12. The *mechanical energy* (ME) of a system is the sum of its kinetic and potential energies.

Law 5.2.13 (Conservation of Mechanical Energy)**.** In the absence of non-conservative forces, the total mechanical energy of an isolated system remains constant.

$$(E_k + E_p)_i = (E_k + E_p)_f.$$

Due to the law of conservation of energy, we have the following (more general) result, which accounts for the presence of dissipative forces:

$$(E_k + E_p)_i + W_d = (E_k + E_p)_f.$$

Note that W_d is negative since the forces are dissipative and hence act in a direction opposite to the displacement.

5.3 Power

Definition 5.3.1. *Power* (P) is the work done (or energy converted) per unit time.

$$P = \frac{\mathrm{d}W}{\mathrm{d}t}.$$

Power is a scalar quantity. Its SI unit is the watt (W).

Proposition 5.3.2. If a constant force F is acting on an object moving at velocity v, the power delivered to the object is given by

$$P = Fv.$$

Proof. By the definition of power,

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = \frac{\mathrm{d}(Fs)}{\mathrm{d}t} = F\frac{\mathrm{d}s}{\mathrm{d}t} = Fv.$$

□

The efficiency of a machine can be expressed in terms of power:

$$\eta = \frac{\text{useful power output}}{\text{power input}}.$$

6 Circular Motion

6.1 Angular Displacement and Velocity

Definition 6.1.1. For a particle moving over an arc of a circle, its *angular displacement* (θ) is the directed angle subtended by the arc at its centre.

Its SI unit is the radian (rad).

Equivalently, the angular displacement θ due to an arc of length s and radius r is given by

$$\theta = \frac{s}{r}.$$

Definition 6.1.2. *Angular velocity* (ω) is the rate of change of angular displacement.

$$\omega = \frac{\mathrm{d}\theta}{\mathrm{d}t}.$$

Its SI unit is radian per second (rad s^{-1}).

Proposition 6.1.3. Assuming the radius of the arc r is constant, the linear velocity v and the angular velocity ω are related by

$$v = r\omega.$$

Proof. We have

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}(r\theta)}{\mathrm{d}t} = r\frac{\mathrm{d}\theta}{\mathrm{d}t} = r\omega.$$

□

The direction of the linear velocity is tangential to the circular path.

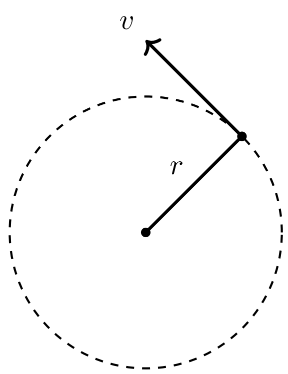

Figure 6.1: v is tangent to the circular path.

6.2 Uniform Circular Motion

Definition 6.2.1. When the angular velocity of a body is constant, the body is said to be moving in *uniform circular motion.*

In uniform circular motion, we simply have

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

Definition 6.2.2. A body moving in a circular path with constant speed is said to be in *uniform circular motion.*

6.2.1 Centripetal Acceleration

Definition 6.2.3. *Centripetal acceleration* (a) is the acceleration of a body in uniform circular motion.

Its SI unit is metres per second squared (m s^{-2}).

Since the magnitude of the body's velocity is constant, the centripetal acceleration only changes the direction of the velocity. Hence, it is perpendicular to the direction of travel, i.e. it points towards the centre of the circular path.

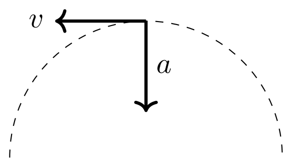

Figure 6.2: The centripetal acceleration a of a particle in uniform circular motion.

Proposition 6.2.4. The centripetal acceleration a of a body in uniform circular motion is given by

$$a = \frac{v^2}{r} = r\omega^2 = v\omega.$$

Proof. Imagine an object steadily traversing a circle of radius r centred on the origin. Its position can be represented by a vector of constant length that changes angle. The total distance covered in one cycle is $2\pi r$. This is also the accumulated amount by which the position has changed.

Now consider the velocity vector of this object: it can also be represented by a vector of constant length that steadily changes direction. This vector has magnitude v, so the accumulated change in velocity is $2\pi v$.

The magnitude of acceleration is then

$$a = \frac{\text{change in velocity}}{\text{time elapsed}} = \frac{\text{change in velocity}}{\text{total distance travelled/speed}} = \frac{2\pi v}{2\pi r/v} = \frac{v^2}{r}.$$

□

6.2.2 Centripetal Force

Definition 6.2.5. *Centripetal force* (F) is the force producing the centripetal acceleration.

Proposition 6.2.6. The centripetal force is given by

$$F = m\frac{v^2}{r} = mr\omega^2 = mv\omega.$$

Proof. From Newton's second law, we know that for an object with constant mass m, we have $F = ma$. But $a = v^2/r = r\omega^2 = v\omega$, so

$$F = m\frac{v^2}{r} = mr\omega^2 = mv\omega.$$

□

7 Gravitational Field

7.1 Gravitational Force and Field

Definition 7.1.1. A *gravitational field* is a region of space within which a mass experiences a gravitational force.

A gravitational field can be represented by directed lines of force.

Law 7.1.2 (Newton's Law of Gravitation)**.** Two point masses attract each other with a force that is proportional to the product of their masses and inversely proportional to the square of their separation. Mathematically,

$$F = \frac{Gm_1m_2}{r^2},$$

where $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2} is the constant of proportionality known as the *gravitational constant*.

Note that point masses have non-zero mass and no volume. If two objects are placed sufficiently far apart such that their dimensions become negligible compared to the distance separating them, the two objects can be considered point masses.

Definition 7.1.3. The *gravitational field strength* (g) at a point is the gravitational force per unit mass exerted on a small test mass placed at that point.

$$g = \frac{F}{m}.$$

g is a vector quantity. It is in the same direction as the gravitational force. Its SI unit is newtons per kilogram (N kg^{-1}).

Proposition 7.1.4. The gravitational field strength g at a point a distance r measured from a point mass M is given by

$$g = \frac{GM}{r^2}.$$

Proof. Consider two point masses M and m separated by a distance r. By Newton's law of gravitation, the magnitude of the attractive gravitational force F acting on the mass m due to the gravitational field of M is given by

$$F = \frac{GMm}{r^2}.$$

Based on the definition of gravitational field strength, the magnitude of g at the point where m is situated is hence

$$g = F/m = \frac{GM}{r^2}.$$

□

7.2 Gravitational Potential Energy

Definition 7.2.1. The *gravitational potential energy* (GPE, U) of a mass at a point is the work done on the mass in moving it from infinity to that point.

Proposition 7.2.2. The gravitational potential energy of a mass m at a distance r_0 from a mass M is given by

$$U = -\frac{GMm}{r_0}.$$

Proof. Recall that

$$F = -\frac{\mathrm{d}U}{\mathrm{d}r}.$$

Hence,

$$U = -\int_{\infty}^{r_0} F\,\mathrm{d}r = -\int_{\infty}^{r_0} \frac{GMm}{r^2}\,\mathrm{d}r = -\frac{GMm}{r_0}.$$

□

Definition 7.2.3. The *gravitational potential* (ϕ) at a point is the work done per unit mass in bringing a small test mass from infinity to that point. Mathematically,

$$\phi = \frac{U}{m}.$$

Gravitational potential is a scalar quantity. Its SI unit is joules per kilogram (J kg^{-1}).

Proposition 7.2.4. The potential at a distance r away from a point mass M is given by

$$\phi = -\frac{GM}{r}.$$

Proof. Recall that $U = -GMm/r$. By the definition of gravitational potential,

$$\phi = \frac{U}{m} = -\frac{GM}{r}.$$

□

Definition 7.2.5. The *potential gradient* in a gravitational field is the change in gravitational potential per unit displacement in the direction of the field. Mathematically,

$$\text{potential gradient} = \frac{\mathrm{d}\phi}{\mathrm{d}r}.$$

Potential gradient is a vector quantity. Its SI unit is joules per kilogram per metre (J kg^{-1} m^{-1}).

Proposition 7.2.6. The field strength g of a gravitational field at a point is numerically equal but opposite in direction to the potential gradient at that point. Mathematically,

$$g = -\frac{\mathrm{d}\phi}{\mathrm{d}r}.$$

Proof. By the definition of g and ϕ, we see that

$$g = \frac{F}{m} = -\frac{1}{m}\frac{\mathrm{d}U}{\mathrm{d}r} = -\frac{\mathrm{d}(U/m)}{\mathrm{d}r} = -\frac{\mathrm{d}\phi}{\mathrm{d}r}.$$

□

The below figure summarizes the relationships between F, U, g and ϕ.

Vectors

$$\begin{tikzcd}
F \arrow[r, "\div m"] & g \arrow[d, "-\int \mathrm{d}r"] \\
U \arrow[u, "-\frac{\mathrm{d}}{\mathrm{d}r}"] & \phi \arrow[l, "\times m"]
\end{tikzcd}$$

Scalars

Figure 7.1: A summary of gravitational fields.

7.3 Orbital Motion

7.3.1 Circular Orbits

Law 7.3.1. For an object in a circular orbit of radius r, the square of its orbital period T is proportional to the cube of the radius.

Proof. Suppose object A (of mass m) is in circular orbit around object B (of mass M). Let T be the orbital period and let r be the distance between A and B.

A is in circular motion. The gravitational force acting on A by B provides the centripetal force. Hence,

$$\frac{GMm}{r^2} = mr\omega^2.$$

Since $\omega = 2\pi/T$, we obtain

$$\frac{GMm}{r^2} = \frac{4\pi^2 mr}{T^2} \implies T^2 = \frac{4\pi^2}{GM} r^3,$$

so $T^2 \propto r^3$. □

Proposition 7.3.2. The kinetic energy of an orbiting mass m is given by

$$E_k = \frac{GMm}{2r}.$$

Proof. Like above, we observe that the gravitational force on the mass provides the centripetal force, so

$$\frac{GMm}{r^2} = \frac{mv^2}{r} \implies E_k = \frac{1}{2}mv^2 = \frac{GMm}{2r}.$$

□

Corollary 7.3.3. The mechanical energy of an orbiting mass m is given by

$$E_{\text{total}} = -\frac{GMm}{2r} = -E_k.$$

Proof. Recall that the mechanical energy is the sum of the mass's KE and GPE. Hence,

$$E_{\text{total}} = E_k + E_p = \frac{GMm}{2r} + \left(-\frac{GMm}{r}\right) = -\frac{GMm}{2r}.$$

□

Paradoxically, if the total energy of the orbiting mass increases, its kinetic energy must *decrease*. However, the process of increasing the total energy is typically started by *increasing* its speed! This paradox arises because $E_k \propto 1/r$; any increase in total must necessarily increase r, which results in a decrease in kinetic energy, and vice versa.

7.3.2 Geostationary Orbits

Definition 7.3.4. A *geostationary orbit* is a circular orbit around the Earth on which an orbiting mass would appear stationary to an observer on the Earth's surface.

For an object to be in geostationary orbit, it must satisfy the following conditions:

- it must be in equatorial orbit (vertically above the equator);
- it must have an orbital period of 24 hours; and
- it must move from west to east.

Part III

Thermal Physics

8 Temperature and Ideal Gases

8.1 Thermal Equilibrium

Definition 8.1.1. The *temperature* of a region is a measure of the average kinetic energy of the molecules in the region.

Definition 8.1.2. *Heat* is the flow of energy between regions of different temperatures.

Definition 8.1.3. Two regions are in *thermal contact* if heat can flow between them (i.e. by conduction, convection, or radiation).

When two regions of different temperatures are in thermal contact, heat will flow from the region with higher temperature to the region with lower temperature. Eventually, the temperatures of the regions will become equal and there will be no further heat transfer.

Definition 8.1.4. Two regions are in *thermal equilibrium* if there is no net heat transfer between them.

Law 8.1.5 (Zeroth Law of Thermodynamics)**.** If bodies A and B are separately in thermal equilibrium with body C, then A and B are also in thermal equilibrium with each other.

8.1.1 Thermometer

Definition 8.1.6. A *thermometer* is an instrument that measures the temperature of a system in a quantitative way.

Definition 8.1.7. A *thermometric property* is a physical property of a system that varies regularly with the system's temperature.

Examples of thermometric properties include

- the volume of a fixed mass of liquid,
- the pressure of a fixed mass of gas at constant volume,
- the resistance of a metal, and
- the electromotive force produced between junctions of dissimilar metals that are at different temperatures.

Definition 8.1.8. An *empirical temperature scale* is a scale of temperature based on a thermometric property that varies linearly with temperature.

Because the thermometric property is linearly related to temperature is linear, we only require two fixed points are usually to establish a scale. Typically, we choose

- the *ice point*, where pure ice can exist in equilibrium with pure water at 1 atm, and
- the *steam point*, where pure water can exist in equilibrium with its pure vapour at 1 atm.

Proposition 8.1.9. Let θ represent temperature and let $x(\theta)$ be a thermometric property that varies linearly with θ. Given two fixed points $(\theta_1, x(\theta_1))$ and $(\theta_2, x(\theta_2))$, we have

$$\theta = \frac{\theta_2 - \theta_1}{x(\theta_2) - x(\theta_2)} \left(x(\theta) - x(\theta_1)\right) + \theta_1.$$

Proof. Since $x(\theta)$ is linear, by the point-slope formula,

$$x(\theta) - x(\theta_1) = \frac{x(\theta_2) - x(\theta_1)}{\theta_2 - \theta_1} \left(\theta - \theta_1\right).$$

Solving for θ yields the claim. □

8.1.2 Kelvin Scale

Definition 8.1.10. *Absolute zero* is the temperature at which all substances have a minimum internal energy.

A temperature scale having absolute zero as its zero point is called an *absolute temperature scale*. Such scales do not depend on the thermometric property of any particular substance.

Definition 8.1.11. The *Kelvin* (K) scale is an absolute temperature scale whose fixed points are absolute zero and the triple point of water.

The triple point of water is the particular temperature and pressure (273.16 K, 4.58 mmHg) at which the three states of water can co-exist in equilibrium.

Temperatures on the Kelvin scale (T / K) and the Celsius scale (θ / °C) are related by

$$T \ / \ \mathrm{K} = \theta \ / \ {}^\circ\mathrm{C} + 273.15.$$

8.2 Ideal Gases

8.2.1 Mole

Definition 8.2.1. The *mole* (mol) is the unit of the amount of substance. One mole of a substance contains as many particles of that substance as there are atoms of carbon in 12 grams of carbon-12.

Definition 8.2.2. *Avogadro's constant* ($N_A = 6.02 \times 10^{23}$ mol^{-1}) is the number of atoms in 12 grams of carbon-12.

Equivalently, one mole of any substance contains 6.02×10^{23} particles.

8.2.2 Equation of State

At low pressures and high temperatures, all gases have a very simple relationship between their pressure p, volume V and temperature T:

$$pV = nRT,$$

where n is the amount of gas in moles and R (= 8.31 J mol^{-1} K^{-1}) is the molar gas constant. This equation is known as the *equation of state of an ideal gas*.

Definition 8.2.3. An *ideal gas* is a hypothetical gas that obeys the equation of state $pV = nRT$ perfectly at all pressure p, volume V, amount of substance n, and temperature T.

If we know that the number of molecules, we can use the equivalent formula

$$pV = NkT,$$

where N $(= nN_A)$ represents the number of gas particles and k $(= 1.38 \times 10^{-23}\ \mathrm{J\ K^{-1}})$ is known as *Boltzmann's constant*.

8.3 Kinetic Theory of Gases

The kinetic theory of gases is a simplified model of the forces associated with the random and continuous motion of gas molecules. It is a microscopic view from which the macroscopic properties of the gas (such as its pressure, volume and temperature), could be explained, derived and/or predicted.

There are six assumptions made by the theory:

1. All gases consist of a very large number of particles.
2. The particles behave as if they are hard, perfectly elastic, identical spheres.
3. There are no forces of attraction or repulsion between particles unless they are in collision with each other or with the walls of the containing vessel.
4. The particles are in constant, random motion and obey Newton's laws of motion.
5. The total volume of particles is negligible compared to the volume of the containing vessel.
6. The time of collisions is negligible compared to the time between collisions.

Under the kinetic theory of gases, the pressure exerted by a gas is the result of the gas molecules colliding with the walls of the container. Each time a gas molecule collides with and rebounds from a wall, it undergoes a change of momentum, indicating a force is exerted on the molecule by the wall (Newton's second law) and an equal but opposite force on the wall by the molecule (Newton's third law). Continuous random collisions by all molecules of the gas would average out to a steady force on the walls. As pressure is force per unit area of contact, the resulting force on the walls of the container is the pressure exerted by the gas.

Definition 8.3.1. Consider N molecules in a gas with respective molecule speeds c_1, c_2, …, c_N. The *mean square speed* ($\langle c^2 \rangle$) of the molecules is the mean of the square of the speeds.

$$\langle c^2 \rangle = \frac{c_1^2 + c_2^2 + \cdots + c_N^2}{N}.$$

The *root-mean-square speed* (c_{rms}) of the molecules is

$$c_{\mathrm{rms}} = \sqrt{\langle c^2 \rangle} = \sqrt{\frac{c_1^2 + c_2^2 + \cdots + c_N^2}{N}}.$$

Proposition 8.3.2. An ideal gas obeys the relationship

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

Proof. Consider a cubical box with side length l containing N particles, each of mass m.

By considering the elastic collision of one particle with a wall, we note that the change in momentum of the particle due to the collision is

$$\Delta p_{\text{particle}} = m\Delta v = m\left((-c_x) - (c_x)\right) = -2mc_x,$$

where c_x is the x-component of the particle's velocity (up to sign convention). Conservation of linear momentum tells us that the wall experiences a change in momentum

$$\Delta p_{\text{wall}} = -\Delta p_{\text{particle}} = 2mc_x.$$

The next time the same particle collides with the same wall, the particle would have travelled a total distance of $2l$ in the same direction. Hence, the time between collisions is

$$\Delta t = \frac{2l}{c_x}.$$

Therefore, the rate of change of momentum of the wall due to a single particle is

$$\frac{\Delta p_{\text{wall}}}{\Delta t} = \frac{2mc_x}{2l/c_x} = \frac{mc_x^2}{l}.$$

According to Newton's second law, the total force acting on the wall is the rate of change of momentum of the wall due to all N particles:

$$F_{\text{wall}} = \sum_{i=1}^{N} \frac{m(c_x)_i^2}{l} = \frac{Nm}{l}\sum_{i=1}^{N} \frac{(c_x)_i^2}{N} = \frac{Nm}{l}\left\langle c_x^2 \right\rangle.$$

The pressure on the wall is force acting per unit area, hence

$$p = \frac{F_{\text{wall}}}{l^2} = \frac{Nm}{l^3}\left\langle c_x^2 \right\rangle = \frac{Nm}{V}\left\langle c_x^2 \right\rangle.$$

By Pythagoras' Theorem, we note that

$$\left\langle c^2 \right\rangle = \left\langle c_x^2 \right\rangle + \left\langle c_y^2 \right\rangle + \left\langle c_z^2 \right\rangle.$$

Because N is large, and the particles are moving randomly (with no preference in any direction), we have $\left\langle c_x^2 \right\rangle = \left\langle c_y^2 \right\rangle = \left\langle c_z^2 \right\rangle$, so

$$\left\langle c^2 \right\rangle = 3\left\langle c_x^2 \right\rangle.$$

It follows that

$$pV = \frac{1}{3}Nm\left\langle c^2 \right\rangle.$$

□

Corollary 8.3.3. The average kinetic energy of an ideal gas particle is $3kT/2$.

Proof. Equating the above result with the equation of state, we see that

$$pV = \frac{1}{3}Nm\left\langle c^2 \right\rangle = NkT \implies \frac{1}{2}m\left\langle c^2 \right\rangle = \frac{3}{2}kT.$$

□

Corollary 8.3.4. The pressure p of an ideal gas of density ρ is given by

$$p = \frac{1}{3}\rho\left\langle c^2 \right\rangle.$$

Proof. Since Nm is the mass of the gas, we have that $\rho = Nm/V$, so

$$p = \frac{1}{3}\frac{Nm}{V}\left\langle c^2 \right\rangle = \frac{1}{3}\rho\left\langle c^2 \right\rangle.$$

□

9 First Law of Thermodynamics

9.1 Specific Heat Capacity and Specific Latent Heat

9.1.1 Specific Heat Capacity

Definition 9.1.1. The *specific heat capacity* (c) of a substance is defined as the heat per unit mass required to cause a unit change in its temperature, without going through a change in state. Mathematically,

$$c = \frac{Q}{m\Delta T},$$

where Q is the heat supplied, m is the mass of the substance, and ΔT is the change in temperature of the substance.

The SI unit of specific heat capacity is J kg^{-1} K^{-1}.

Definition 9.1.2. The *heat capacity* (C) of a substance is defined as the heat required to cause a unit change in its temperature, without going through a change in state. Mathematically,

$$C = \frac{Q}{\Delta T} = mc.$$

The SI unit of heat capacity is J K^{-1}.

9.1.2 Specific Latent Heat

Definition 9.1.3. The *specific latent heat* (L) of a substance is the amount of heat per unit mass required to change its state without changing its temperature. Mathematically,

$$L = \frac{Q}{m}.$$

The SI unit of specific latent heat is J kg^{-1}.

For the change of state between solid and liquid, it is known as the specific latent heat of *fusion*, and it occurs at the substance's melting point.

For the change of state between liquid and gas, it is known as the specific latent heat of *vaporization*, and it occurs at the substance's boiling point.

9.2 Internal Energy

A body possesses energy due to its physical state and its state properties (e.g. pressure, temperature), because in changing from one state to another, it can do work (e.g. causing itself or other bodies to move and/or gain energy).

Definition 9.2.1. The *internal energy* (U) of a body is its ability to do work due to its state. It can be expressed as the sum of a random distribution of kinetic and potential energies associated with the molecules of a system. Mathematically,

$$U = E_k + E_p,$$

where E_k is the internal kinetic energy and E_p is the internal potential energy.

The internal kinetic energy of a substance is the total energy of the atoms and molecules in the substance due to their translational, rotational and vibrational motion. It depends on the temperature of the substance.

The internal potential energy of a substance is the total energy of the atoms and molecules in the substance due to intermolecular forces. It depends on the separation of the atoms and molecules.

Note that the internal energy of a system depends only on the state that it is in, and not on how it reached that state.

Proposition 9.2.2. The internal energy U of an ideal gas with N molecules is given by

$$U = \frac{3}{2}NkT = \frac{3}{2}pV = \frac{3}{2}nRT.$$

Proof. An ideal gas is assumed to have negligible intermolecular forces of attraction, hence $E_p = 0$ and the internal energy is simply $U = E_k$. Recall that the kinetic energy of a single particle is $3kT/2$, so the total kinetic energy of the gas is

$$U = E_k = \frac{3}{2}NkT.$$

□

This implies that at absolute zero ($T = 0$), the internal energy of an ideal gas is 0.

9.2.1 Melting and Boiling

From the kinetic theory, the mean random kinetic energy of the molecules in a substance is directly proportional to its temperature.

In melting and boiling, the heat supplied is used by the molecules to do work against the intermolecular attractions, allowing them to move freely within the substance, increasing E_p. Meanwhile, no work is done to increase the mean kinetic energy of the molecules, hence temperature and E_k remain the same.

In boiling, however, the heat supplied is also used to do work against atmospheric pressure. Given further that the molecules now much further apart (as compared to boiling), the specific latent heat of vaporization is hence higher than the specific latent heat of fusion.

9.3 First Law of Thermodynamics

Law 9.3.1. The increase in the internal energy of a system is the sum of the heat supplied to the system and the work done on the system.

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

Note that if the heat is supplied by the system, Q would be negative. Likewise, if work is done by the system, W would be negative.

The work done on the system depends on the expansion/contraction of the system. Recall that

$$W = -\int_{V_1}^{V_2} p \, \mathrm{d}V.$$

If $V_2 < V_1$, volume decreases and W is positive, hence work is done on the system. If $V_2 > V_1$, volume increases and W is negative, hence work is done by the system.

9.3.1 Thermodynamic Processes

Definition 9.3.2. A *thermodynamic process* refers to a change in the state of the system. The succession of states through which the system passes is the *path* of the process.

The state of a system can be represented by a point on a p-V graph.
There are four common thermodynamic processes for an ideal gas:

Definition 9.3.3. In an *isochoric* process, the volume of the system is constant.

Since volume is constant, no work is done. Hence, the first law of thermodynamics simplifies to $\Delta U = Q$.

Definition 9.3.4. In an *isobaric* process, pressure is constant.

Since pressure is constant, the work done is simply $W = p\Delta V$.

Definition 9.3.5. In an *isothermal* process, temperature is constant.

From the equation of state $pV = nRT$, we see that pV is also constant, so

$$\Delta U = \frac{3}{2}p_i V_i - \frac{3}{2}p_f V_f = 0.$$

From this, we deduce that $Q = -W$.

Definition 9.3.6. In an *adiabatic* process, there is no heat exchange between the system and the surroundings ($Q = 0$).

From the first law of thermodynamics, we have $\Delta U = W$.

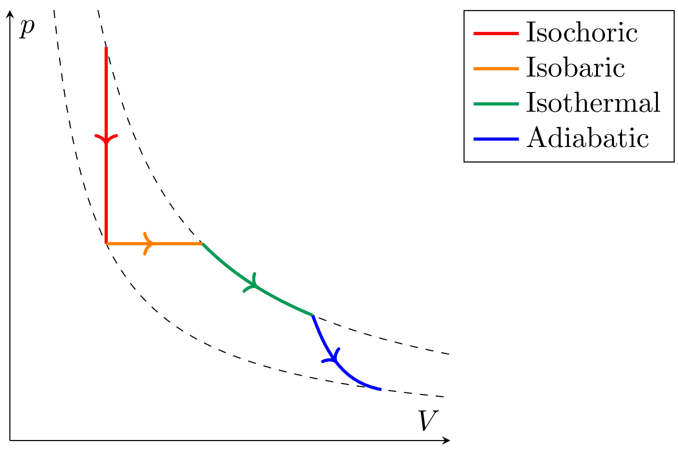

Figure 9.1: The four common thermodynamic processes.

Part IV

Oscillation and Waves

10 Oscillations

Definition 10.0.1. An *oscillation* is a periodic to-and-fro motion of an object between two limits.

Definition 10.0.2. The *amplitude* (x_0) is the maximum displacement of the oscillating object from its equilibrium position.

Amplitude is a scalar quantity. Its SI unit is the metre.

Definition 10.0.3. The *displacement* (x) is the distance of the oscillating object from its equilibrium position in a stated direction.

Displacement is a vector quantity. Its SI unit is the metre.

Definition 10.0.4. The *period* (t) is the time taken for the oscillating object to undergo one complete oscillation.

Period is a scalar quantity. Its SI unit is the second.

Definition 10.0.5. The *frequency* (f) is the number of complete oscillations per unit time.

Frequency is a scalar quantity. Its SI unit is the Hertz (Hz).
Frequency is the reciprocal of period.

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

Definition 10.0.6. The *angular frequency* (ω) is the angle per unit time, with one completely oscillation represented as 2π rad.

Angular frequency is a scalar quantity. Its SI unit is radians per second (rad s^{-1}).
Angular frequency is related to frequency and period via

$$\omega = 2\pi f = \frac{2\pi}{T}.$$

Definition 10.0.7. *Phase* is an angle which gives a measure of the fraction of a cycle that has been completed by the oscillating object.

Phase is a scalar quantity. Its SI unit is the radian (rad).

Definition 10.0.8. The *phase difference* between two oscillations is a measure of how much one oscillation is out of step with another.

Phase difference is a scalar quantity. Its SI unit is the radian (rad).

10.1 Simple Harmonic Motion

There are three types of oscillations: free, damped, and forced. We look at free oscillations first.

Definition 10.1.1. *Free oscillations* refer to oscillations with no external force acting, and so have constant energy and amplitude.

The displacement-time graph of a free oscillation is sinusoidal. Mathematically, we can describe the displacement x with respect to time t with the general equations

$$x = x_0 \sin(\omega t + \phi) \quad \text{or} \quad x = x_0 \cos(\omega t + \phi)\,.$$

Here, the phase of the oscillating object is $\omega t + \phi$.

Definition 10.1.2. *Simple harmonic motion* (SHM) is the oscillatory motion of a particle whose acceleration is proportional but opposite in direction to its displacement.

$$a \propto -x.$$

We typically define ω^2 to be the constant of proportionality:

$$a = -\omega^2 x.$$

This is known as the *defining equation* for simple harmonic motion.

Proposition 10.1.3. The equations $x = x_0 \sin(\omega t + \phi)$ and $x = x_0 \cos(\omega t + \phi)$ satisfy the defining equation $a = -\omega^2 x$.

Proof. Since the two expressions are phase shifts of each other, it suffices to show that either one satisfies the defining equation. We work with $x = x_0 \sin(\omega t + \phi)$. Indeed, one can verify that the acceleration a is given by

$$a = \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = -x_0 \omega^2 \sin(\omega t + \phi) = -\omega^2 x,$$

so $x = x_0 \sin(\omega t + \phi)$ is a solution to the defining equation for simple harmonic motion. □

Proposition 10.1.4. Given $x = x_0 \sin(\omega t + \phi)$, the following equations hold:

$$v = x_0 \omega \cos(\omega t + \phi) = \pm\omega\sqrt{x_0^2 - x^2} \quad \text{and} \quad a = -x_0 \omega^2 \sin(\omega t + \phi) = -\omega^2 w.$$

Proof. The velocity-time, acceleration-time and acceleration-displacement equations can easily be derived by repeatedly differentiating $x = x_0 \sin(\omega t + \phi)$ with respect to time. We now show that the velocity-displacement equation $v = \pm\omega\sqrt{x_0^2 - x^2}$ holds.

First, observe that

$$v^2 = x_0^2 \omega^2 \cos^2(\omega t + \phi) \quad \text{and} \quad x^2 \omega^2 = x_0^2 \omega^2 \sin^2(\omega t + \phi)\,.$$

Summing the two expressions, we obtain

$$v^2 + x^2\omega^2 = x_0^2 \omega^2 \implies v^2 = \omega^2 \left(x_0^2 - x^2\right).$$

Taking roots yields

$$v = \pm\omega\sqrt{x_0^2 - x^2}.$$

□

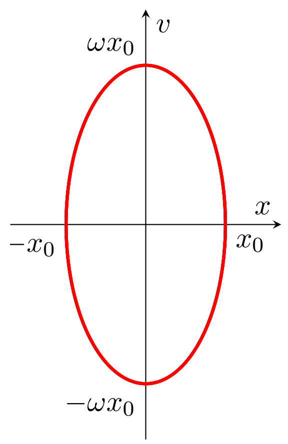

Figure 10.1: A graph of velocity against displacement.

10.1.1 Energy in Simple Harmonic Motion

Proposition 10.1.5. The kinetic energy E_k of a body of mass m undergoing simple harmonic motion is given by

$$E_k = \frac{1}{2}m\omega^2\left(x_0^2 - x^2\right).$$

Proof. Recall that $v^2 = \omega^2\left(x_0^2 - x^2\right)$, so

$$E_k = \frac{1}{2}mv^2 = \frac{1}{2}m\omega^2\left(x_0^2 - x^2\right).$$

□

Corollary 10.1.6. The maximum kinetic energy is

$$\max E_k = \frac{1}{2}m\omega^2 x_0^2$$

and occurs when $x = 0$.

Since mechanical energy is conserved, we know that

$$E_{\text{total}} = E_k + E_p.$$

When the kinetic energy of an oscillating object is maximum, its potential energy must be minimum. If we define $\min E_p = 0$, then

$$E_{\text{total}} = \max E_k + \min E_p = \frac{1}{2}m\omega^2 x_0^2.$$

Proposition 10.1.7. The potential energy E_p of a body undergoing simple harmonic motion is given by $\frac{1}{2}m\omega^2x^2$.

Proof. We have

$$E_p = E_{\text{total}} - E_k = \frac{1}{2}m\omega^2 x_0^2 - \frac{1}{2}m\omega^2\left(x_0^2 - x^2\right) = \frac{1}{2}m\omega^2 x^2.$$

□

10.2 Damped Oscillations

Definition 10.2.1. *Damped oscillation* occurs when there is continuous dissipation of energy to the surroundings, resulting in a decrease in the amplitude of the oscillating system over time.

There are three degrees of damping.

- With *light damping*, the object undergoes a number of complete oscillations with the amplitude decreasing exponentially with time.
- With *critical damping*, a displaced system returns to rest at its equilibrium position in the shortest possible time without oscillating.
- With *heavy damping*, a displaced system takes a long time to return to its equilibrium position without oscillating.

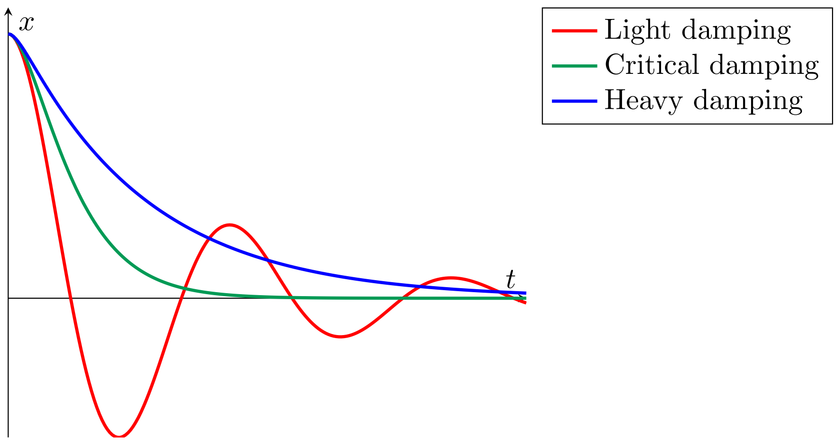

Figure 10.2: The displacement-time graphs of the three degrees of damping.

10.3 Forced Oscillations

Definition 10.3.1. *Forced oscillations* occur when an external periodic force continuously supplies energy to an oscillating system to counteract energy lost through damping, thereby maintaining a constant amplitude of oscillation.

When a system oscillates in the absence of any external force, it does so at its own *natural frequency* (f_0). Whatever its natural frequency, a forced oscillation takes on the frequency of the driving oscillations, called the *driving frequency*.

The amplitude of a forced oscillation varies with the driving frequency f.

Definition 10.3.2. *Resonance* refers to the phenomenon that the oscillation amplitude reaches a maximum when the driving frequency is equal to the natural frequency of the forced oscillation.

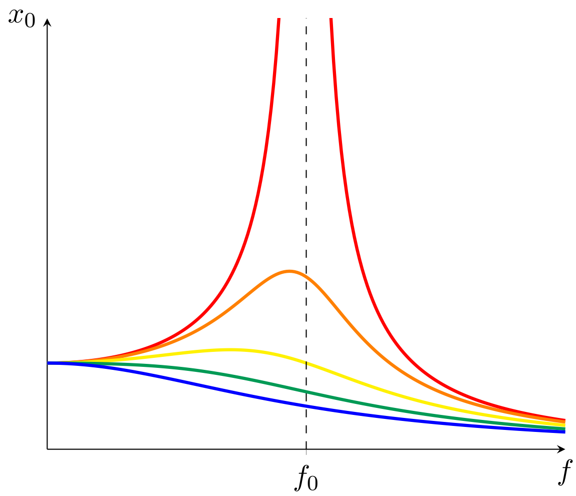

Figure 10.3: Frequency response graph for varying degrees of damping. Red represents no damping, blue represents heavy damping.

As damping increases, we observe that

- the response is less sharp (lower and flatter peak, and smaller amplitudes at all frequencies), and
- the maximum amplitude (i.e. resonance) occurs when the driving frequency is slightly less than the natural frequency.

11 Wave Motion

Definition 11.0.1. *Wave motion* refers to the transfer of energy by propagation of oscillations without net movement of the medium.

Propagation of oscillations in physical media (e.g. air, water) are called mechanical waves. Propagation of oscillations in electromagnetic fields are called electromagnetic waves.

Definition 11.0.2. A *wavelength* (λ) is the minimum distance between any two points of the wave with the same phase at the same instant.

The SI unit of wavelength is the metre (m).

Definition 11.0.3. The *wave speed* (v) is the speed with which energy is transmitted by a wave.

The SI unit of wave speed is metre per second (m s^{-1}).

Proposition 11.0.4. The wave speed v is related to the wavelength λ and frequency f by

$$v = f\lambda.$$

Proof. By definition,

$$v = \frac{\text{distance travelled in 1 oscillation}}{\text{time taken for 1 oscillation}} = \frac{\lambda}{T} = f\lambda.$$

□

Definition 11.0.5. A *wave font* is an imaginary line or surface joining points which are in phase.

Wave fronts are usually drawn one wavelength apart and are often thought to represent wave crests. All the points on a wave front have the same distance from the source of the wave.

Definition 11.0.6. A *ray* is the direction in which the energy of a wave is travelling.

Rays are always perpendicular to wave fronts.

Definition 11.0.7. The *phase difference* ($\Delta\phi$) between two waves at a point or between 2 points on a wave is the difference in the phases of their oscillation cycle expressed as an angle, where 2π represents one cycle.

The phase difference between two waves of the same frequency at a point is equivalent to the time difference for the waves to be at the same oscillation phase, expressed as an angle, where 2π represents one period.

$$\Delta\phi = \frac{\Delta t}{T} \cdot 2\pi.$$

The phase difference between two points on a wave is equivalent to the distance between the points expressed as an angle, where 2π represents one wavelength.

$$\Delta f = \frac{\Delta x}{\lambda} \cdot 2\pi.$$

11.1 Progressive Waves

Definition 11.1.1. A *progressive wave* is a wave in which energy is carried by means of vibration or oscillation within the waves, without transporting matter.

Progressive waves can be further classified as either transverse of longitudinal.

Definition 11.1.2. A *transverse wave* is a wave whose vibrations are perpendicular to the direction of transfer of energy of the wave.

Definition 11.1.3. A *longitudinal wave* is a wave whose vibrations are parallel to the direction of transfer of energy of the wave.

All electromagnetic waves are transverse.

Definition 11.1.4. The *intensity* (I) of a wave is the rate of energy transmitted per unit area perpendicular to the wave propagation.

The SI unit of intensity is watt per square metre (W m^{-2}).

Proposition 11.1.5. For a fixed frequency f, the intensity I of a wave is proportional to the square of its amplitude A.

Proof. Recall from simple harmonic motion that the total energy associated with an oscillation is given by

$$E = \frac{1}{2}m\omega^2 A^2 = \frac{1}{2}m\left(2\pi f\right)^2 A^2 = \underbrace{\left(2\pi^2 m f^2\right)}_{\text{constant}} A^2.$$

Since $I \propto E$ and $E \propto A^2$ (for constant f), so $I \propto A^2$. □

Proposition 11.1.6. In 3D, the intensity I at a point located a distance r from a point source which emits energy with power P is given by

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

Proof. The area of all points a distance r away from the point source is $4\pi r^2$. Hence, by the definition of I,

$$I = \frac{\text{energy/time}}{\text{area}} = \frac{\text{power}}{\text{area}} = \frac{P}{4\pi r^2}.$$

□

In 2D, we have the analogous result:

Proposition 11.1.7. In 2D, the intensity[1] I at a point located a distance r from a point source which emits energy with power P is given by

$$I = \frac{P}{2\pi r}.$$

[1]Note that in 2D, we modify the definition of intensity to "rate of energy transmitted per unit length perpendicular to the wave propagation."

11.2 Polarization

A transverse wave could have a mixture of oscillations in an infinite number of directions, in the plane normal to the direction of energy transfer. This is known as an *unpolarized wave*.

Definition 11.2.1. *Polarization* is the process by which a wave's oscillations are made to occur in one direction only, in the plane normal to the direction of energy transfer.

Polarization is a phenomenon associated only with transverse waves (e.g. light). This is because in longitudinal waves, the medium only oscillates in one direction (parallel to the direction of energy transfer).

Proposition 11.2.2. The final intensity I of an unpolarized light wave (of initial intensity I_0) after passing through a polarizer is given by

$$I = \frac{1}{2}I_0.$$

Proof. The unpolarized light wave is a random mixture of all states of polarization, and the vertical and horizontal components are, on average, equal. Hence, the polarizer should transmit 50% of all incident light, i.e.

$$I = \frac{1}{2}I_0.$$

□

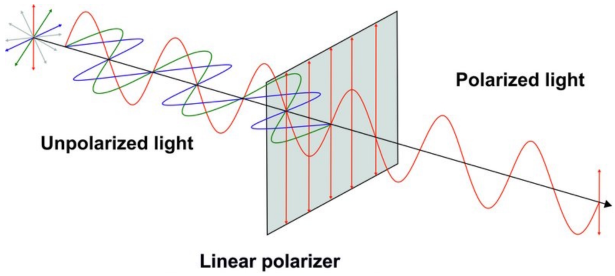

Figure 11.1: Effect of a polarizer on unpolarized light.[2]

Theorem 11.2.3 (Malus' Law)**.** When a polarizer is placed in a polarized beam of light, the intensity I of the light that passes through is given by

$$I = I_0 \cos^2\theta,$$

where I_0 is the initial intensity and θ is the angle between the light's initial polarization direction and the axis of the polarizer.

Proof. Let A be the initial amplitude of the polarized light wave. When the light passes through the new polarizer, only the components that are parallel to the new polarization direction passes through. The resulting amplitude is thus $A\cos\theta$. Since intensity is proportional to amplitude squared, we have

$$I = k\,(A\cos\theta)^2 \qquad \text{and} \qquad I_0 = kA^2$$

for the same constant of proportionality k, so

$$\frac{I}{I_0} = \cos^2\theta \implies I = I_0\cos^2\theta.$$

□

[2]Source: https://www.codixx.com/knowledge-corner/polarization

12 Superposition

12.1 Principle of Superposition

Principle 12.1.1 (Principle of Superposition)**.** When two or more waves of the same type meet, the resultant displacement is the vector sum of the individual displacements of the waves.

Definition 12.1.2. When the phase difference between two waves at a point is an integer multiple of a cycle ($\Delta\phi = 2k\pi$), the waves are in *phase*. When the phase difference is an odd integer multiple of a half cycle ($\Delta\phi = (2k+1)\pi$), the waves are in *anti-phase*.

Definition 12.1.3. *Interference* is the result of superposing two or more waves of the same type.

In particular, when two waves of the same frequency and amplitude superpose in phase, they reinforce each other and the resultant displacement is twice that of each individual wave. This is called *constructive interference*.

In contract, when two waves of the same frequency and amplitude superpose in anti-phase, they cancel each other and the resultant displacement is zero. This is called *destructive interference*.

A result between the two extremes is called a *partial interference*.

Definition 12.1.4. Two waves of the same type and frequency are said to be *coherent* if there is a constant phase difference between them.

Two wave sources are said to be coherent if the waves they produce are coherent.

Definition 12.1.5. The *path difference* to a particular point from two wave sources is the difference in path lengths from the two sources to that point.

12.2 Stationary Waves

Definition 12.2.1. A wave is said to be *stationary* (or *standing*) if energy is not transferred, but stored in the oscillations of the medium.

When two progressive waves of the same type of equal amplitude, equal frequency, and equal speed travelling in opposite directions meet, the resultant wave is stationary.

Definition 12.2.2. A *node* is a point on a standing wave where the displacement is always zero.

Definition 12.2.3. An *anti-node* is a point on a standing wave with maximum amplitude.

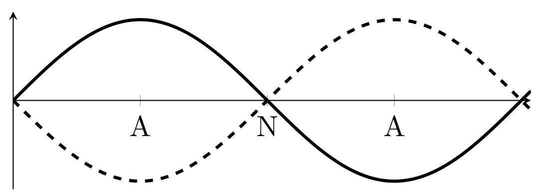

Figure 12.1: Nodes and anti-nodes on a standing wave.

Nodes occur when the two underlying progressive waves meet in anti-phase, while anti-nodes occur when the two underlying progressive waves meet in phase.

The distance between two adjacent nodes or two adjacent anti-nodes is $\lambda/2$, where λ is the wavelength of the underlying progressive waves as well as that of the resultant stationary wave.

In a stationary wave, all particles within a loop vibrate in phase, but have a phase difference of π rad with particles in an adjacent loop.

12.2.1 Stationary Waves in Stretched Strings

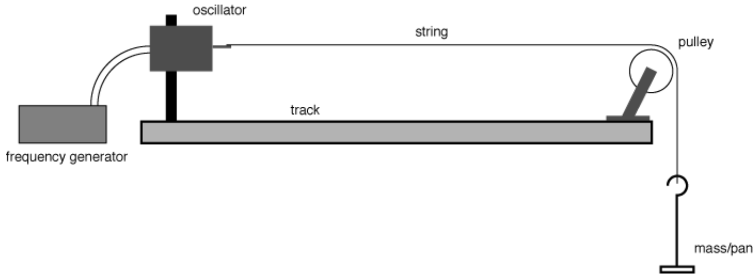

Figure 12.2: Generating stationary waves in a stretched string.[1]

The set up above shows how stationary waves can be produced in strings.

A string of length L is held taut by connecting a mass on one end through a pulley and the other end to an oscillator. A frequency generator is used to vary the frequency of the oscillation.

The progressive wave produced by the oscillator moves towards the pulley with a constant velocity v. The wave is reflected at the pulley and travels backwards. The incident and reflected waves have the same speed, frequency and amplitude, moving in opposite directions.

For a stationary wave to form, the wave must have a node at both ends of the string. That is, there must be an integer number of loops on the length of the string. The resultant stationary wave is known as a *harmonic*.

Proposition 12.2.4. In a string, the nth harmonic occurs when n loops fit exactly into the length of the string.

$$L = n\left(\frac{\lambda}{2}\right) \qquad \text{for } n = 1, 2, 3, \ldots.$$

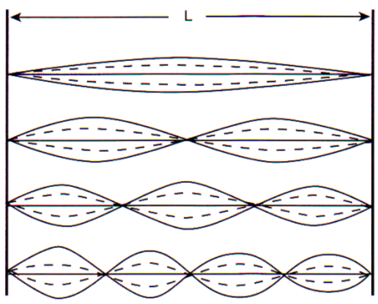

Figure 12.3: The first four harmonics ($n = 1, 2, 3, 4$).[2]

[1]Source: https://physicslabs.ccnysites.cuny.edu/labs/208/208-vibrating-strings/vibrating-strings.php

[2]Source: https://standingwavesch4.wordpress.com/harmonics/

12.2.2 Stationary Waves in Air Columns

Stationary waves can be formed in both closed pipes (open at only one end) and open pipes (open at both ends).

When air is blown into the open end of the pipe, a wave travels from the open end towards the closed end. The wave is reflected when it hits the wall of the closed end of the pipe. A stationary wave forms only if the closed ends are nodes and the open ends are anti-nodes.

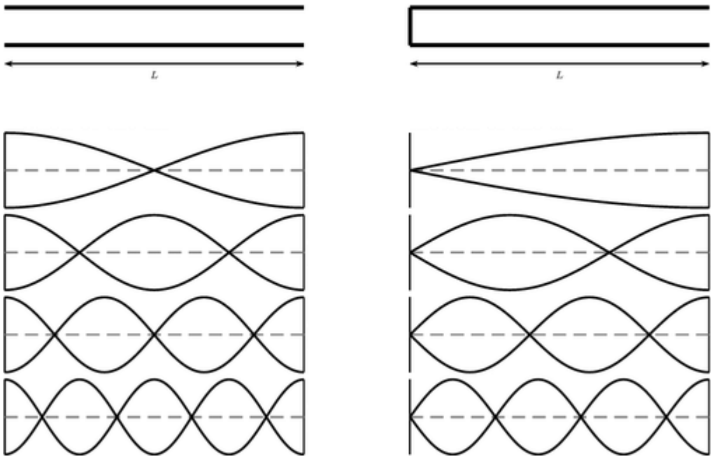

Figure 12.4: The first four harmonics in an open and closed pipe.[3]

In practice, the anti-node at the open end occurs slightly outside the pipe. Hence, there is an *end correction* (c) that needs to be considered when calculating the wavelength or frequency of the wave.

Proposition 12.2.5. In an open pipe, the nth harmonic occurs when n A-N-A cycles fit exactly into the length of the pipe (accounting for end corrections).

$$L = n\left(\frac{\lambda}{2}\right) + 2c \quad \text{for } n = 1, 2, 3, \ldots.$$

Proposition 12.2.6. In a closed pipe, the nth harmonic occurs when n A-N-A cycles fit exactly into *twice* the length of the pipe (accounting for end corrections).

$$2L = n\left(\frac{\lambda}{2}\right) + 2c \quad \text{for } n = 1, 3, 5, \ldots.$$

Note that closed pipes only have odd harmonics.

[3] Source: https://link.springer.com/article/10.1007/s00233-019-10059-4

12.3 Diffraction

When waves encounter an obstacle or an opening, they will spread out after passing through an opening or around the edge of the obstacle.

Definition 12.3.1. *Diffraction* is the spreading of a wave as it passes through a gap or around an obstacle.

The spreading of the wavefronts becomes more pronounced when the size of the slit or *aperture* (a) is of the same order as the wavelength (λ), i.e. $a \approxeq \lambda$. However, it is negligible when the width is large in comparison to the wavelength, i.e. $a \gg \lambda$.

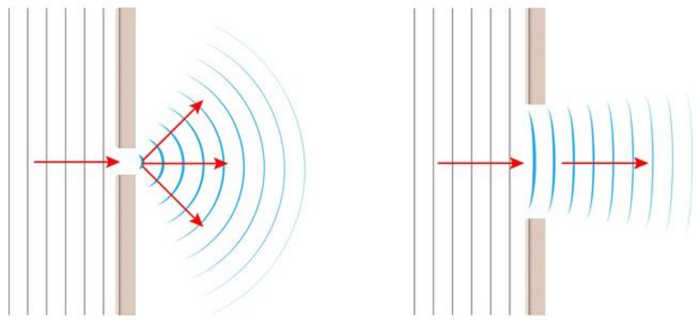

Figure 12.5: The significance of the spreading depends on the relative size of the aperture.[4]

12.3.1 Single Slit Diffraction

The figure below shows the experimental arrangement for observing the diffraction of light through a narrow slit of width b. A viewing screen is placed at a distance L $(\gg b)$ behind the slit.

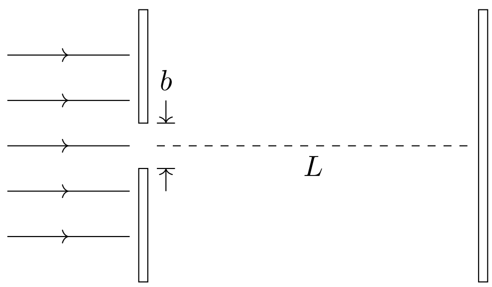

Figure 12.6: The set-up for a single slit experiment.

The light pattern on the screen consists of a central maximum flanked by a series of weaker secondary maxima and dark fringes. It is also observed that the central maximum is significantly broader and brighter than the secondary maxima.

Proposition 12.3.2. Points of zero intensity (i.e. dark fringes) occur at angles θ satisfying

$$b \sin\theta = m\lambda \quad \text{for } m = \pm 1, \pm 2, \pm 3, \ldots.$$

[4]Source: https://www.alamy.com/wave-diffraction-wave-impinges-on-a-narrow-and-a-broad-slit-comparison-of-large-and-small-opening-waves-spread-out-beyond-the-gap-vector-diagram-image473787292.html

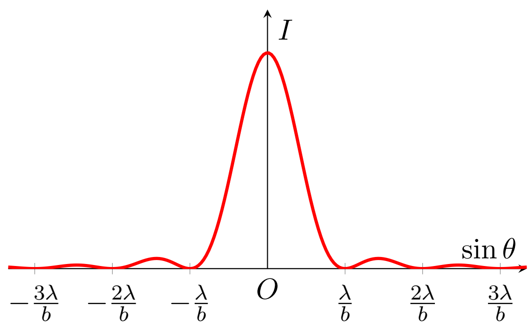

Figure 12.7: The intensity distribution for a single slit diffraction.

12.3.2 Rayleigh's Criterion

The *resolving power* of an optical instrument refers to its ability to distinguish between the images of relatively close objects.

Resolving power is often limited by diffraction phenomenon. The images from two separate points cannot be distinguished if the central maxima of their diffraction pattern overlaps. They are considered to be *just resolved* if the central maxima of their images coincides with the first minima of the other.

Proposition 12.3.3 (Rayleigh Criterion)**.** For the two points to be just resolved, the central maximum of one must lie on the first minimum of the other, i.e.

$$b\sin\theta = \lambda,$$

where θ is the angle between the two sources.

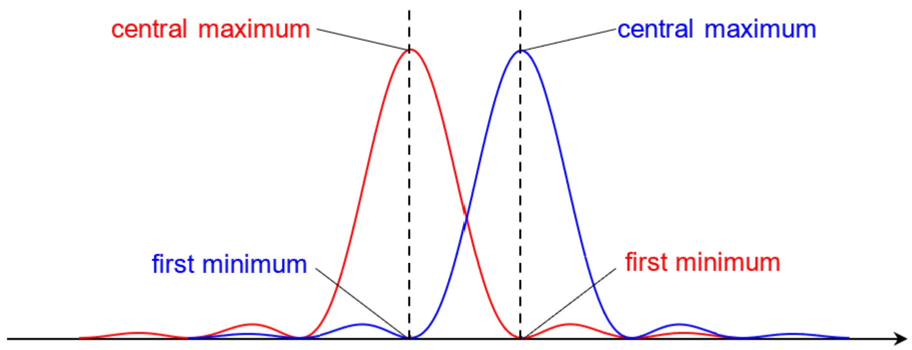

Figure 12.8: A sketch of Rayleigh's criterion.[5]

[5]Source: https://xmphysics.com/2023/01/02/10-5-5-rayleighs-criterion-2/

12.4 Two-Source Interference

A stationary wave is an example of two-source interference of waves travelling in opposite directions. The nodes and anti-nodes are examples of destructive and constructive interference respectively.

12.4.1 Water Waves

When two dippers are attached to the oscillator of a ripple tank, two sets of circular water waves are generated and superpose. The resulting interference pattern is shown below.

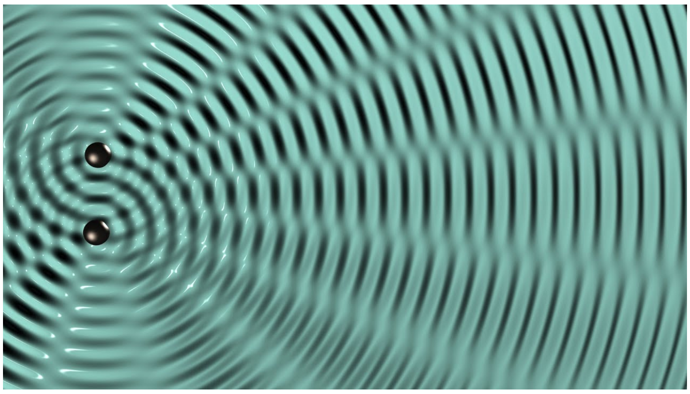

Figure 12.9: Interference pattern produced by water waves.[6]

Note that the two dippers are coherent sources since they are attached to the same oscillator.

Constructive interference occurs at points where the waves meet in phase (i.e. the path differences from the two sources are an integer number of wavelengths). These points form lines called *anti-nodal lines*.

Destructive interference occurs at points where the waves meet in anti-phase (i.e. the path differences are a half-integer number of wavelengths). These points form lines called *nodal lines*.

12.4.2 Observable Interference

Conditions for interference patterns to be observable, besides the obvious condition that the waves must superpose, include the following:

- The waves must be coherent. For light waves, this condition implies that they must be monochromatic.
- The waves must have the same amplitude.
- For transverse waves, they must be either unpolarized or polarized in the same plane.

Note that two separate light sources are not coherent.

12.4.3 Young's Double-Slit Experiment

A *Young's double-slit experiment* involves passing light from a monochromatic source through a single slit followed by a double slit, as shown in the following diagram.

Diffraction at the single slit produces a point light source. Diffraction at the double slits produces two coherent light sources.

[6]Source: https://www.youtube.com/watch?v=fjaPGkOX-wo

The lights from the double splits superpose on the screen with alternate constructive and destructive interference, resulting in a fringe pattern of bright and dark lines on the screen.

The bright lines on the screen are regions where the light arrive in phase (i.e. the path differences from the double slits are an integer number of wavelengths), and constructive interference occurs. The dark lines are regions where the light arrive in anti-phase (i.e. the path differences are a half-integer number of wavelengths), and destructive interference occurs.

Proposition 12.4.1. The wavelength λ of the monochromatic light is related to the double-slit separation a, fringe spacing x and screen distance D, by the expression

$$\lambda = \frac{ax}{D},$$

provided $a \ll D$.

Effect of Interference and Diffraction on Intensity Distribution

The effects of interference *redistribute* the energy into bright and dark fringes. The total energy is conserved.

Suppose that the original light wave has intensity I_0 and amplitude A, where $I_0 = kA^2$ for some constant of proportionality k. At the bright fringes, the waves constructively interfere, so the intensity due to interference is

$$I_{\text{interference}} = k\left(A + A\right)^2 = 4kA^2 = 4I_0.$$

At the dark fringes, amplitude is zero, so intensity is also zero.

Further, as the width of the slits are comparable to the wavelength of light, diffraction occurs at the two slits.

The resultant intensity distribution is due to both the effect of interference and diffraction, i.e.

$$I = I_{\text{interference}} \cdot I_{\text{diffraction}},$$

as shown in the figure below.

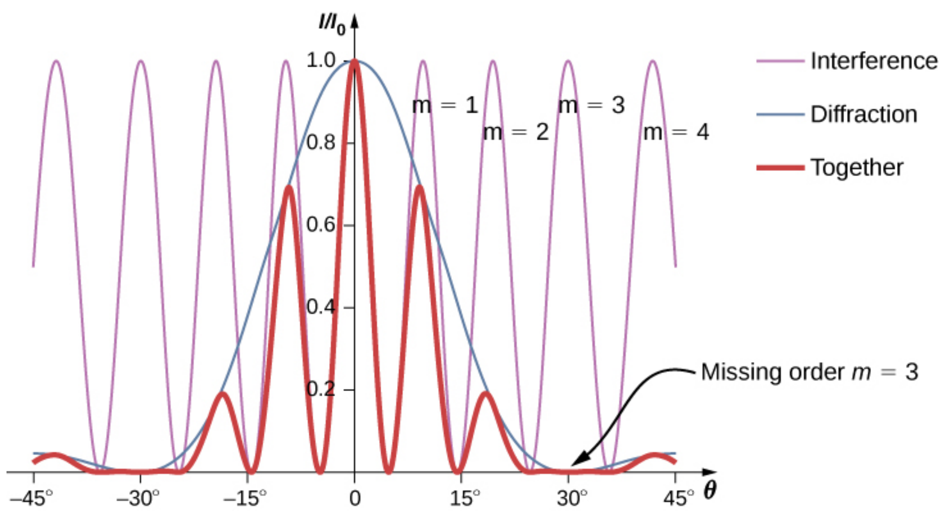

Figure 12.10: The intensity distribution of a double-slit experiment.[7]

[7]Source: https://pressbooks.online.ucf.edu/osuniversityphysics3/chapter/double-slit-diffraction/

Note that some maxima of interference may be missing, such as the maxima of order ± 3 in the above figure. This is because at those points, the diffraction effect is at a minimum.

12.5 Diffraction Grating

A *diffraction grating* consists of a large number of fine, equidistant and closely spaced parallel lines of equal width, ruled on glass or polished metal.

When monochromatic light passes through a grating, diffraction occurs at each of the slits. The waves emerging from the slits are coherent and interference occurs in the region where the waves overlap.

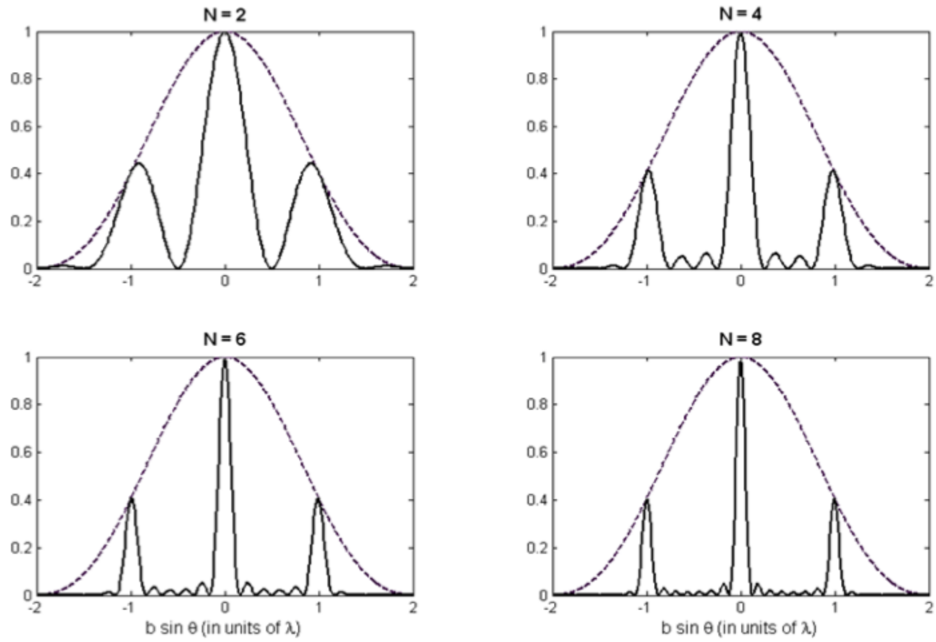

Figure 12.11: Intensity distribution of diffraction grating for varying number of slits N.[8]

Proposition 12.5.1. The nth order maximum occur at angles θ satisfying

$$d \sin\theta = n\lambda \quad \text{for } n = 1, 2, 3, \ldots,$$

where d is the slit separation.

[8]Source: https://www.physicsbootcamp.org/section-diffraction-grating-light.html

Part V

Electricity and Magnetism

13 Electric Fields

13.1 Electric Force and Field

Definition 13.1.1. The *electric charge* is a property of matter that gives rise to electric forces.

The charge on electrons are designated as negative and that on protons are designated as positive. Like charges repel while unlike charges attract.

The SI unit of electric charge is the coulomb (C), which is defined as ampere second (A s).

Definition 13.1.2. A *charged body* refers to a body that has gained or lost some electrons.

A body that has gained electrons has a (net) negative charge while one that has lost electrons has a (net) positive charge.

13.1.1 Electric Field

Definition 13.1.3. An *electric field* is a region of space within which an electric charge experiences an electric force.

An electric field can be represented by directed lines of force. Electric field lines indicate the paths that a small positive test charge would take if it is placed in the field. They always begin on a positive charge and end on a negative charge. Electric field lines do not intersect.

The resultant field of two or more charges can be represented by directed lines which represent the vector sum of their individual fields.

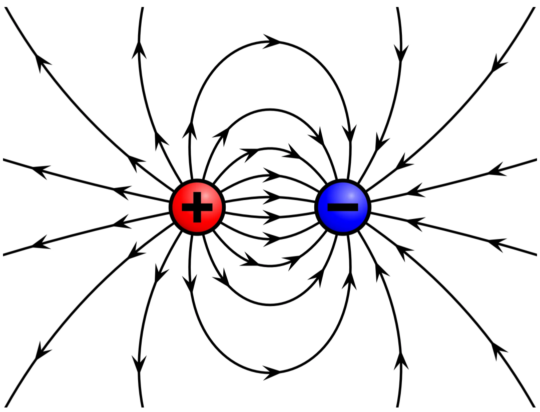

Figure 13.1: The electric field lines of a field consisting of a positive and a negative charge of equal magnitude.[1]

On a conducting body, any net charge would reside entirely on its surface. There are no field lines inside the body. Field lines just outside a charged conducting body are always perpendicular to the body's surface.

On an isolated body, there would be more charge per unit area on sharper surfaces and less on flatter surfaces.

[1]Source: https://www.faqs.com.pk/what-is-electric-field/

Law 13.1.4 (Coulomb's Law). Two point charges exert a force on each other that is proportional to the product of their charges and inversely proportional to the square of their separation.

Mathematically,

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2},$$

where ε_0 ($= 8.85 \times 10^{-12}$ m^{-3} kg^{-1} s^4 A^2) is the *permittivity of free space.*

Definition 13.1.5. The *electric field strength* (E) at a point in an electric field is the electric force per unit positive charge exerted on a small test charge placed at that point.

$$E = \frac{F}{q}.$$

Electric field strength is a vector quantity. Its SI unit is newton per coulomb (N C^{-1}) or volt per metre (V m^{-1}).

Proposition 13.1.6. The electric field strength E due to a point charge Q at a distance r from the charge is given by

$$E = \frac{Q}{4\pi\varepsilon_0 r^2}.$$

Proof. From Coulomb's law, we see that

$$E = \frac{F}{q} = \frac{Q}{4\pi\varepsilon_0 r^2}.$$

□

13.2 Electric Potential Energy

Definition 13.2.1. The *electric potential energy* (U) at a point is the work done in bringing a small test charge from infinity to that point.

Proposition 13.2.2. In an electric field with point charges $Q_1, \ldots, Q_n$ at distances $r_1, \ldots, r_n$ from a point charge q, the electrical potential energy of q is given by

Proof. Recall that $F = -\mathrm{d}U/\mathrm{d}r$, so

$$U_i = -\int_\infty^{r_i} F \,\mathrm{d}r = -\int_\infty^{r_i} \frac{Q_i q}{4\pi\varepsilon_0 r^2} \,\mathrm{d}r = \frac{q}{4\pi\varepsilon_0} \int_\infty^{r_i} -\frac{Q_i}{r^2} \,\mathrm{d}r = \frac{q}{4\pi\varepsilon_0} \frac{Q_i}{r_i}.$$

Adding all n contributions together, we get

$$U = \frac{q}{4\pi\varepsilon_0} \sum_{i=1}^{n} \frac{Q_i}{r_i}$$

as desired. □

Definition 13.2.3. The *electric potential* (V) at a point is the work done per unit positive charge in bringing a small test charge from infinity to that point. Mathematically, we have

$$V = \frac{U}{q}.$$

Electric potential is scalar quantity. Its SI unit is the volt (V) or joule per coulomb (J C^{-1}).

Proposition 13.2.4. The electric potential V due to a point charge Q at a distance r_0 from the charge is given by

$$V = \frac{Q}{4\pi\varepsilon_0 r}.$$

Proof. By definition, we have

$$V = \frac{U}{q} = \frac{Q}{4\pi\varepsilon_0 r}.$$

□

Electric potential may be represented by lines of equal potential. The net work done to move a charge between any two points on the same equipotential line is zero.

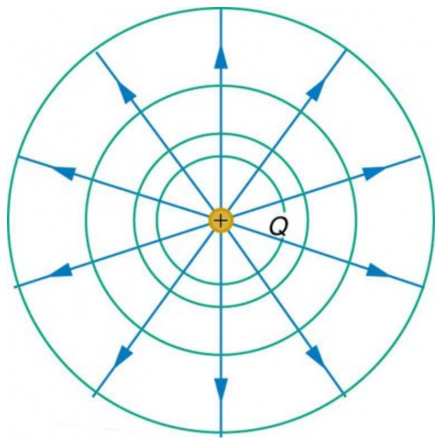

Figure 13.2: Equipotential lines near a point charge.[2]

Note that equipotential lines are perpendicular to the field lines.

Definition 13.2.5. The *potential gradient* in an electric field is the change in electric potential per unit displacement in the direction of the field. Mathematically,

$$\text{potential gradient} = \frac{\mathrm{d}V}{\mathrm{d}r}.$$

Potential gradient is a vector quantity. Its SI unit is joules per kilogram per metre (J kg^{-1} m^{-1}).

Proposition 13.2.6. The electric field strength E at a point is numerically equal but opposite in direction to the potential gradient at that point. Mathematically,

$$E = -\frac{\mathrm{d}V}{\mathrm{d}r}.$$

Proof. By the definition of E and V, we see that

$$E = \frac{F}{q} = -\frac{1}{q}\frac{\mathrm{d}U}{\mathrm{d}r} = -\frac{\mathrm{d}U/q}{\mathrm{d}r} = -\frac{\mathrm{d}V}{\mathrm{d}r}.$$

□

The below figure summarizes the relationships between F, U, E and V. Observe the parallels with gravitational fields.

[2]Source: https://courses.lumenlearning.com/suny-physics/chapter/19-4-equipotential-lines/

Vectors

$$\begin{tikzcd}
F \arrow[r, "\div q"] & E \arrow[d, "-\int \mathrm{d}r"] \\
U \arrow[u, "-\frac{\mathrm{d}}{\mathrm{d}r}"] & V \arrow[l, "\times q"]
\end{tikzcd}$$

Scalars

Figure 13.3: A summary of electric fields.

13.3 Uniform Electric Fields

The electric field between a pair of charged parallel plates with a small separation is almost uniform near the centre. The field becomes more uniform as the plate separation decreases and/or the area of the plates increases.

Proposition 13.3.1. The uniform field strength E between a pair of parallel plates with potential difference V and separation d is given by

$$E = \frac{V}{d}.$$

Proof. Observe that

$$qV = U = W = Fd = qEd \implies E = \frac{V}{d}.$$

□

Within a uniform electric field, where the field strength E is the same at all points, a charged particle will experience a constant electric force in the direction of the field if the charge is negative, and in the opposite direction if it is negative. The motion of charged particles in a uniform electric field is therefore one of uniform acceleration in a direction parallel to the field, and the equations of motion can be used to analyse its trajectory.

14 Current of Electricity

14.1 Electric Current

Definition 14.1.1. *Electric current* (I) is the rate of flow of charge.

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t}.$$

The SI unit of electric current is the ampere (A).

Definition 14.1.2. An *electric conductor* is a material that contains mobile charge carriers (e.g. electrons). Else, it is an *electric insulator*.

Electric current is conventionally taken to be in the direction of the flow of positively-charged particles, even though in many conductors, negatively charged electrons or ions are the charge carriers.

Proposition 14.1.3. The current I in a wire is given by

$$I = nAvq,$$

where n is the number of charge carriers per volume, A is the cross-sectional area of the wire, v is the drift velocity of the charge carriers, and q is the charge on each charge carrier.

Proof. Let V be the volume of the wire and l the length of the wire, so $V = Al$. Let N be the total number of charge carriers. Then the total charge Q is given by

$$Q = \text{charge per charge carrier} \times \text{total number of charge carriers} = (q)\,(nV) = qnAl.$$

Assuming constant charge, we have from the definition of I that

$$I = \frac{Q}{t} = \frac{qnAl}{t} = qnA\left(\frac{l}{t}\right) = qnAv.$$

$\square$

14.2 Potential Difference and Electromotive Force

Definition 14.2.1. The *potential difference* (V, p.d.) or *voltage* between two points in a circuit is the work done per unit charge when electrical energy is transferred to non-electrical energy when the charge passes from one point to the other.

$$V = \frac{W}{Q}.$$

The SI unit of potential difference is the volt (V).

Definition 14.2.2. The electromotive force (E, e.m.f.) of a source is the work done per unit charge when non-electrical energy is transferred into electrical energy when the charge is moved round a complete circuit.

The SI unit of electromotive force is the volt (V).

Note that the electromotive force is a source of energy and will always exist, even when a current is not flowing in the circuit. However, potential difference exists only when current is flowing in the circuit.

14.3 Resistance

Definition 14.3.1. The *resistance* (R) of a conductor is the ratio of the potential difference across the conductor to the current passing through it.

$$R = \frac{V}{I}.$$

The SI unit of resistance is the ohm (Ω).

Law 14.3.2 (Ohm's Law)**.** A current I flowing through a conductor is directly proportional to the potential difference V applied across the conductor, provided that physical conditions (like temperature, stress of the material, etc.) remain constant.

$$I \propto V.$$

Definition 14.3.3. The *resistivity* (ρ) of a material is a measure of its resistance to current flowing through it. It relates the resistance R of the material to its length l and cross-sectional area A via

$$R = \rho \frac{l}{A}.$$

The SI unit of resistivity is ohm metre (Ω m).

14.3.1 I-V Characteristics

The I-V characteristic (or I-V curve) shows how the current I flowing through a device varies with the voltage V across it.

When the current passing through a material increases, the temperature of the material is likely to increase. Two main changes occur at the molecular level which affects the resistance of the material:

- the number of charge carriers per unit volume (n) increases, which reduces the resistance of the material.
- the thermal vibrations of the lattice atoms increases, which increases the number of collisions between the charged carriers and the lattice atoms. This reduces the drift velocity v of the charged carriers, hence increasing the resistance of the material.

Depending on the nature of the material, one effect may dominate, which will determine the variation of I with respect to V.

Ohmic Resistors

An *ohmic resistor* refers to a resistor with a constant resistance.

Note that ohmic resistors obey Ohm's law, where $1/R$ is the constant of proportionality between I and V:

$$I = \frac{V}{R}.$$

The I-V characteristic of an ohmic resistor is a straight line through the origin.

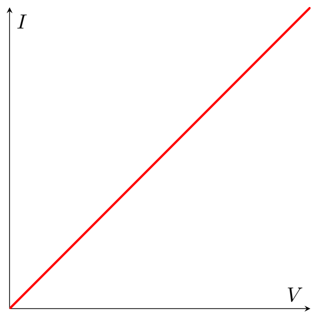

Figure 14.1: The I-V characteristic of an ohmic resistor.

A metallic conductor at constant temperature is an example of an ohmic resistor. At constant temperature, its lattice vibrations remain the same, and so its resistance remains the same.

Filament Lamp

The I-V characteristic of a filament lamp is increasing but concave downwards.

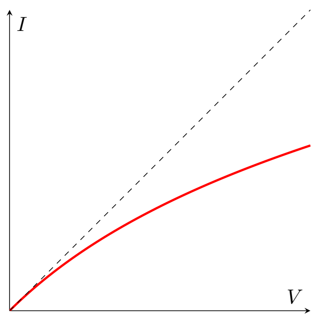

Figure 14.2: The I-V characteristic of a filament lamp.

As temperature increases, the number of electrons in a filament lamp does not vary significantly. However, lattice vibrations increases, which reduces the drift velocity of electrons, hence leading to an increase in resistance.

Semiconductor Diode

A *diode* is a device that has a low resistance in one direction and a very high resistance in the other direction.

If a potential difference is applied across a diode in the direction with low resistance, it is said to be *forward biased*. If a potential difference is applied in the direction with very high resistance, it is said to be *reverse biased*.

The I-V characteristic of a semiconductor diode is increasing and concave upwards when forward biased. When reverse biased, the I-V cure is almost zero, except when the voltage exceeds the breakdown voltage and the current becomes very large.

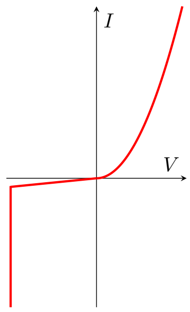

Figure 14.3: The I-V characteristic of a semiconductor diode.

Consider the forward biased region of the I-V graph. As V increases, the temperature of the semiconductor increases, hence electrons have more energy and are more likely to escape from atoms. This increases n significantly. At the same time, there is also an increase in the rate of interaction of electrons with the lattice vibrations. However, the increase in n predominates over the rate of lattice interactions, hence the overall effect is that resistance decreases.

Negative Temperature Coefficient Thermistor

A *thermistor* is a resistor with a resistance that is temperature dependent. A negative temperature coefficient (NTC) thermistor is a thermistor whose resistance decreases with temperature.

The I-V characteristic of an NTC thermistor is increasing and concave upwards.

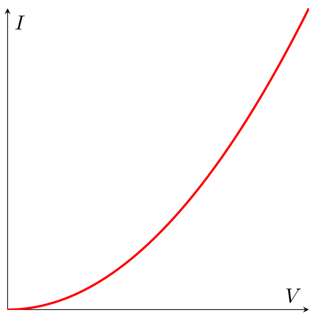

Figure 14.4: The I-V characteristic of a NTC thermistor.

As V increases, the temperature of the NTC thermistor increases, resulting in lower resistance.

14.4 Power

Proposition 14.4.1. The power P dissipated between two points is given by the product of the potential difference V and the current I between the two points, i.e. $P = IV$.

Proof. From the definition of potential difference,

$$V = \frac{W}{Q} = \frac{W/t}{Q/t} = \frac{P}{I},$$

which rearranges to $P = IV$. □

Corollary 14.4.2. The power P is given by

$$P = I^2 R = \frac{V^2}{R}.$$

Proof. Recall that $V = IR$, so

$$P = IV = I\,(IR) = I^2 R \quad \text{and} \quad P = IV = \left(\frac{V}{R}\right) V = \frac{V^2}{R}.$$

□

14.5 Internal Resistance

In practice, whenever there is a current, some of the energy transferred by a source is dissipated in the source itself. The source is said to have *internal resistance* (r).

The internal resistance of a source could be depicted in a circuit diagram as a resistor r connected in series with an ideal source of electromotive force E, connected to an external circuit of combined resistance R.

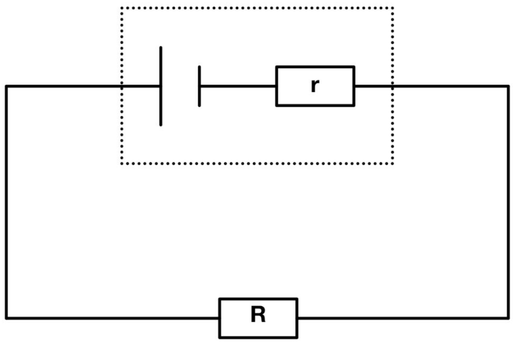

Figure 14.5: A simplified circuit diagram showing the internal resistance of a source.[1]

As the electromotive force is the total energy (per unit charge) available to the circuit and some energy is dissipated due to internal resistance, the potential difference available to the external circuit will be less than the electromotive force.

Proposition 14.5.1. The potential difference V available to the external circuit is given by

$$V = E - Ir.$$

Proof. Conservation of energy implies that the electromotive force E must be equal to the sum of the potential difference across the external and internal resistances, i.e.

$$E = IR + Ir = V + Ir \implies V = E - Ir.$$

□

Various quantities can be expressed in terms of E, r and R.

[1] Source: https://thescienceandmathszone.com/internal-resistance/

Proposition 14.5.2.

- Current $I = \dfrac{E}{R+r}$
- Terminal potential difference $V = IR = \dfrac{RE}{R+r}$.
- Total power dissipated in the complete circuit $P_T = IE = \dfrac{E^2}{R+r}$.
- Power dissipated in the external circuit $P_R = IV = \dfrac{RE^2}{(R+r)^2}$.
- Efficiency $\eta = \dfrac{P_R}{P_T} = \dfrac{R}{R+r}$.

Theorem 14.5.3 (Maximum Power Transfer Theorem)**.** Maximum power is supplied to the external circuit when the resistance of the external circuit is equal to the internal resistance of the source.

Proof. From above, we see that the power dissipated in the external circuit is

$$P_R = \frac{RE^2}{(R+r)^2}.$$

By the AM-GM inequality,

$$(R+r)^2 \geq 4Rr \implies P_R = \frac{RE^2}{(R+r)^2} \leq \frac{RE^2}{4Rr} = \frac{E^2}{4r},$$

which depends only on the source. Equality occurs when $R = r$, as desired. □

15 D.C. Circuits

15.1 Kirchhoff's Laws

Law 15.1.1 (Conservation of Charge)**.** Charge can neither be created nor destroyed.

As a consequence of the conservation of charge, the total charge that enters a junction per unit time must be equal to the total charge that leaves the same junction per unit time. This is better known as Kirchhoff's current law.

Law 15.1.2 (Kirchhoff's Current Law)**.** The sum of currents entering any junction in an electric circuit is always equal to the sum of currents leaving that junction.

Recall that the law of conservation of energy states that energy can neither be created nor destroyed. Therefore, the electrical energy produced by the sources should be equal to the sum of electrical energy consumed by all the components in a circuit. This is known as Kirchhoff's voltage law.

Law 15.1.3 (Kirchhoff's Voltage Law)**.** In any closed loop in an electric circuit, the total electromotive force E supplied is equal to the total potential difference in that loop.

15.1.1 Effective Resistance

Proposition 15.1.4. The effective resistance R of resistors connected in series is the sum of their individual resistances, i.e.

$$R = R_1 + R_2 + \cdots + R_n.$$

Proof. Consider two resistors in series with a potential difference V applied across them. By Kirchhoff's current law, the current passing through each resistor must be equal. Hence,

$$V = IR_1 + IR_2 = I\left(R_1 + R_2\right).$$

If the resistors were replaced by a single resistor of resistance R such that the same current I would flow when the same potential difference V is applied, then

$$V = IR.$$

Equating, we obtain $R = R_1 + R_2$. □

Proposition 15.1.5. The reciprocal of the effective resistance R of resistors connected in parallel is the sum of the reciprocals of their individual resistances, i.e.

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}.$$

Proof. Consider two resistors in parallel with a potential difference V applied across them. By Kirchhoff's voltage law, the potential difference across each resistor must be equal, i.e.

$$V = I_1R_1 = I_2R_2.$$

This implies that the total current I is

$$I = I_1 + I_2 = \frac{V}{R_1} + \frac{V}{R_2} = V\left(\frac{1}{R_1} + \frac{1}{R_2}\right).$$

If the resistors were replaced by a single resistor of resistance R such that the same total current I would flow when the same potential difference V is applied, then

$$V = IR \implies I = \frac{V}{R}.$$

Equating, we obtain

$$R = \frac{1}{R_1} + \frac{1}{R_2}.$$

□

Corollary 15.1.6. For resistors in parallel, the combined resistance is always less than any of the individual resistances.

15.2 Potential Dividers

Definition 15.2.1. A *potential divider* is an arrangement of resistors which is used to obtain a fraction of the potential difference provided by a voltage supply.

The circuit below shows a simple divider which consists of two known resistors of resistances R_1 and R_2 connected in series to a voltage supply of electromotive force E. The potential difference V_{out} across R_2 is then connected to a load.

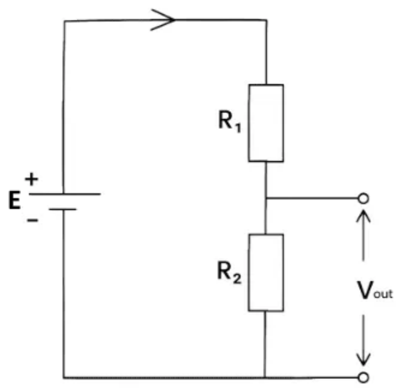

Figure 15.1: A typical circuit diagram for a potential divider.[1]

Proposition 15.2.2. The potential difference V_{out} across R_2 is given by

$$\frac{V_{\text{out}}}{E} = \frac{R_2}{R_1 + R_2}.$$

Proof. The current I passing through the resistors is

$$I = \frac{E}{R_1 + R_2},$$

so the potential difference across R_1 is

$$V_{\text{out}} = IR_2 = \frac{R_2}{R_1 + R_2}E$$

and the desired claim follows. □

[1]Source: https://studymind.co.uk/notes/potential-dividers/

15.2.1 Potentiometers

Ideally, when a potential difference is being measure, no current should be drawn from the circuit involved. However, in practice, most voltmeters draw a small current from the circuit.

To remedy this, a *potentiometer* can be used to compare potential differences without drawing a current from the circuit involved.

Definition 15.2.3. A *potentiometer* is an adjustable potential divider (typically using a sliding contact).

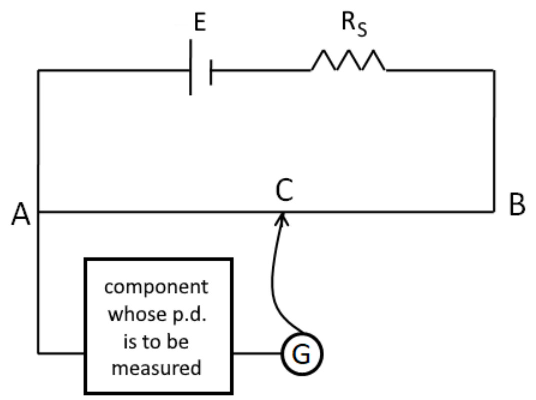

Figure 15.2: A typical potentiometer circuit.

For a potentiometer found in laboratories, a length of resistance wire AB is used, acting as two resistors (AC and CB). An (optional) additional resistor R_S is connected in series with the wire, serving to limit the current that flows through the circuit. The resistance of R_S will affect the potential difference across AB.

A sliding contact is connected to a galvanometer G as shown, and is moved along the wire until a point, C, is found on the wire such that there is no current in the galvanometer (null reading). This point is known as the *null/balance point* and the length AC is known as the *balance length*. At the null point, the potential difference across AC is equal to the potential difference across the component to be measured. Hence, no current will flow through the galvanometer.

Definition 15.2.4. The *potential gradient* (ϕ) of a potentiometer wire is the potential difference per unit length of the potentiometer wire.

The SI unit of the potential gradient is volt per metre (V m^{-1}).

Principle 15.2.5 (Potentiometer Principle)**.** The potential gradient ϕ of a potentiometer wire is constant, i.e.

$$\frac{V_{AB}}{L_{AB}} = \frac{V_{AC}}{L_{AC}} = \phi.$$

Proof. Recall that $R = \rho L/A$, so

$$\phi = \frac{V}{L} = \frac{IR}{L} = \frac{I\rho}{A},$$

which is constant. □

Measuring Electromotive Forces

To measure the electromotive force of a source, the potentiometer circuit is to be set up as shown below.

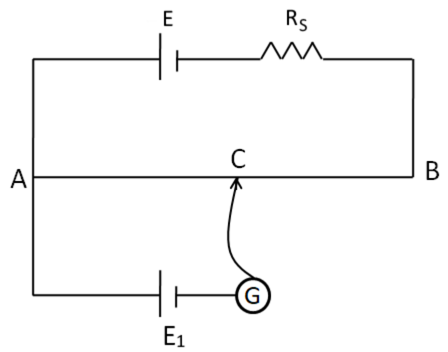

Figure 15.3: The set-up used to measure the electromotive force of a source using the potentiometer principle.

Proposition 15.2.6. The electromotive force of the source in the above figure is given by

$$\frac{E_1}{E} = \frac{L_{AC}}{L_{AB}} \cdot \frac{R}{R + R_s},$$

where R is the resistance of wire AB.

Proof. Let the potential difference across wire AB be V_{AB}. Then

$$\frac{V_{AB}}{E} = \frac{R}{R + R_S}.$$

By the potentiometer principle,

$$\frac{E_1}{L_{AC}} = \frac{V_{AB}}{L_{AB}},$$

so

$$\frac{E_1}{E} = \frac{L_{AC}}{L_{AB}} \cdot \frac{V_{AB}}{E} = \frac{L_{AC}}{L_{AB}} \cdot \frac{R}{R + R_S}.$$

□

Measuring Internal Resistance

To determine the internal resistance of a source, the potentiometer circuit is to be set up as shown below.

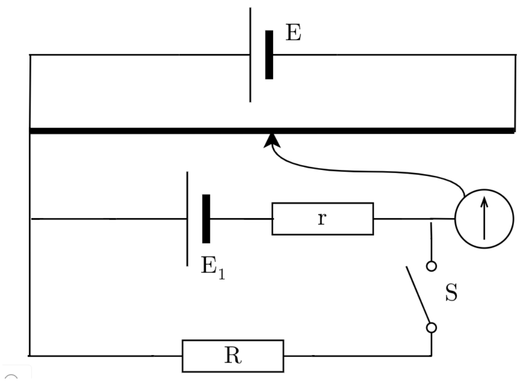

Figure 15.4: The set-up used to measure the internal resistance of a source using the potentiometer principle.

Proposition 15.2.7. The internal resistance of the cell is given by

$$\frac{r}{R} = \frac{L_1}{L_2} - 1,$$

where L_1 and L_2 are the balance lengths when switch S is open and closed respectively.

Proof. When switch S is open, no current passes through the branch circuit, hence the potential difference across the branch circuit comes solely from the electromotive force E_1. Thus,

$$\phi = \frac{E_1}{L_1}.$$

When switch S is closed, we can treat the potentiometer as a potential divider. We previously derived that

$$\frac{V}{E_1} = \frac{R}{R+r},$$

where V is the terminal potential difference. On the other hand,

$$\phi = \frac{V}{L_2}.$$

Hence, by the potentiometer principle,

$$\frac{E_1}{L_1} = \frac{V}{L_2} \implies \frac{R}{R+r} = \frac{V}{E_1} = \frac{L_2}{L_1} \implies \frac{r}{R} = \frac{L_1}{L_2} - 1.$$

□

16 Electromagnetism

Definition 16.0.1. A *magnetic field* is a region of space within which a current-carrying conductor, a moving charge, or a permanent magnet may experience a magnetic force.

A magnetic field can be represented with directed flux lines. Magnetic flux lines are always closed loops that do not start or stop in mid-space. The lines indicate the paths that a magnetic north pole would take if it is placed in the field.

Definition 16.0.2. The *magnetic flux density* or *magnetic field strength* (B) at a point is the magnetic force per unit length of conductor per unit current carried placed perpendicular to the field at that point.

Mathematically,

$$B = \frac{F}{IL\sin\theta},$$

where θ is the angle between B and I.

The SI unit of magnetic flux density is the tesla (T), or newton per metre per ampere (N m^{-1} A^{-1}).

16.1 Magnetic Field due to Currents

In this section, μ_0 is the *permeability of free space*.

16.1.1 Long Straight Wire

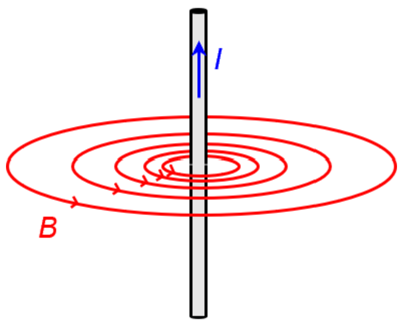

Figure 16.1: The magnetic field lines for a long straight wire.[1]

Proposition 16.1.1. The flux density B at a distance d from a wire is given by

$$B = \mu_0 \frac{I}{2\pi d}.$$

The direction of the flux around a wire can be identified using the right-hand grip rule: when the thumb of the right-hand points in the direction of the conventional current, the curled fingers point in the direction of the magnetic field lines.

[1]Source: https://xmphysics.com/2023/01/10/14-1-1-wires-coils-and-solenoids/

16.1.2 Flat Circular Coil

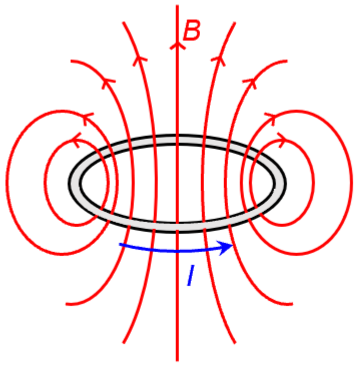

Figure 16.2: The magnetic field lines for a flat circular coil.[1]

Proposition 16.1.2. The flux density B at the centre of a coil with N turns of radius r is given by

$$B = \mu_0 \frac{NI}{2r}.$$

The direction of the flux can be identified using the right-hand grip rule: if the right-hand is gripping the coil so that the fingers curl in the same direction as conventional current, the extended thumb will point in the direction of the field lines inside the coil, i.e. it points to the North pole.

16.1.3 Long Solenoid

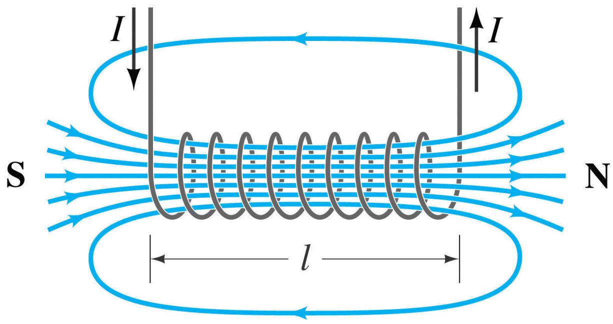

Figure 16.3: The magnetic field lines for a long solenoid.[2]

Proposition 16.1.3. The flux density B at the centre of a long solenoid of length L with N turns is given by

$$B = \mu_0 \frac{NI}{L}.$$

The direction of the flux can be identified using the right-hand grip rule.

The flux density inside a solenoid may be influenced by the presence of a core with a different permeability from that of free space. For instance, with a ferrous ("soft" iron) core, which has a permeability of about $5000\mu_0$, the flux density could be increased by a factor of about 5000.

Solenoids with a core of high permeability are called *electromagnets*.

[2]Source: https://www.miniphysics.com/ss-magnetic-field-due-to-current-in-a-solenoid.html

16.2 The Motor Effect

16.2.1 Current-Carrying Conductor

Proposition 16.2.1. A conductor of length L carrying a current I at an angle θ to a uniform magnetic field of flux density B will experience a force F given by

$$F = BIL\sin\theta.$$

Proof. Recall the definition of magnetic flux:

$$B = \frac{F}{IL\sin\theta} \implies F = BIL\sin\theta.$$

□

The direction of the magnetic force, which is perpendicular to both the field and the current, can be predicted using *Fleming's left-hand rule*: if the thumb and the first two fingers of the left hand are held so that they are mutually at right angles, then they represent force (F), field (B) and current (I) respectively.

Current Balance

The force on a current-carrying conductor can be used to measure the flux density of a magnetic field using a *current balance*.

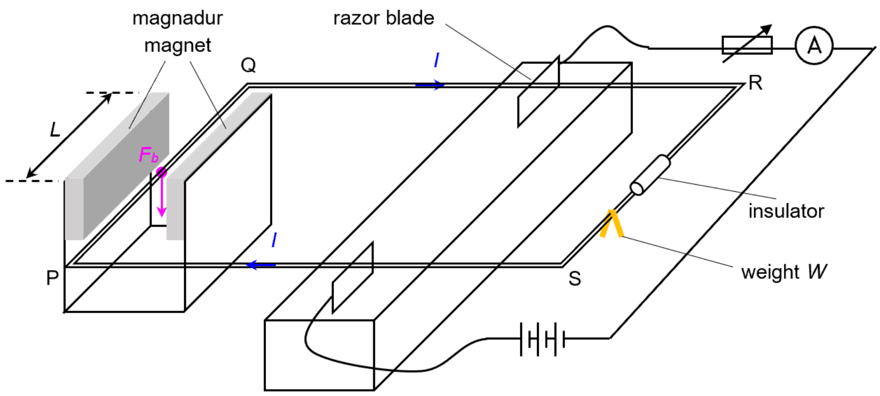

Figure 16.4: An example of a current balance.[3]

For example, in the above set-up, side PQ of a rectangular coil is placed perpendicularly inside the magnetic field generated by two magnets of side length L. The coil is balanced on a pair of razor blades. A weight W is added to the RS side of the coil. A current is passed through the coil and is adjusted until the coil is in equilibrium.

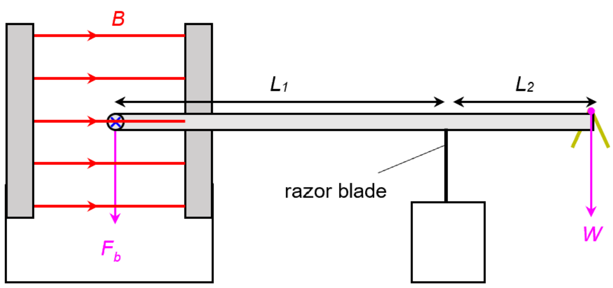

Figure 16.5: The side view of the current balance.[3]

[3]Source: https://xmphysics.com/2023/01/11/14-2-3-current-balance/

The magnitude of B can then be calculated by equation the anti-clockwise moment caused by the magnetic force $F = BIL$, and the clockwise moment caused by the weight W about the razor blade pivot:

$$(BIL)\,L_1 = WL_2 \implies B = \frac{W}{IL}\frac{L_2}{L_1}.$$

16.2.2 Moving Charge

Proposition 16.2.2. A charge Q moving at velocity v at an angle θ to a uniform magnetic field of flux density B will experience a force F given by

$$F = BQv\sin\theta.$$

Proof. Observe that

$$IL = \left(\frac{Q}{t}\right)L = Q\left(\frac{L}{t}\right) = Qv,$$

so

$$F = BIL\sin\theta = BQv\sin\theta.$$

□

Proposition 16.2.3. If a charged particle enters a uniform magnetic field at a right-angle to the flux, its path in the field will be circular with radius

$$r = \frac{mv}{QB},$$

where m is the mass of the particle, v its velocity and Q its charge.[4]

Proof. By Fleming's left-hand rule, the velocity v of the particle is perpendicular to the magnetic field B, thus the magnetic force provides the centripetal force for the particle to move in a circle. Equating the two, we see that

$$BQv\sin 90^\circ = \frac{mv^2}{r} \implies r = \frac{mv}{QB}.$$

□

5

Corollary 16.2.4. The period of revolution of the particle is independent of its speed.

Proof. We have

$$T = \frac{2\pi r}{v} = \frac{2\pi\left(\frac{mv}{QB}\right)}{v} = \frac{2\pi m}{Bq},$$

which is independent of v. □

[5]A similar result holds when the particle enters the field at an angle. In this case, however, the particle will move in a helical path.

Velocity Selector

Uniform magnetic and electric fields that are perpendicular to each other can be used to select charged particles of a particular speed. This arrangement produces forces opposite in directions acting on a beam of charged particles passing through the fields.

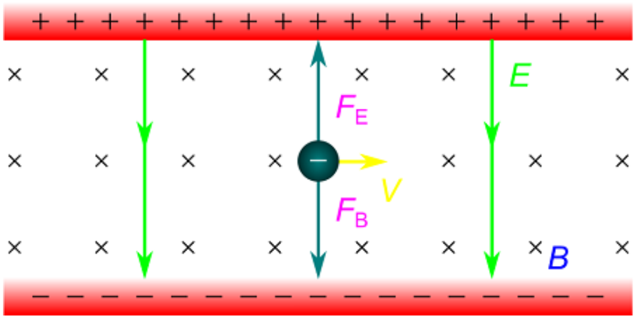

Figure 16.6: An example of a velocity selector. Here, the magnetic field lines are into the page.

A particle of charge Q and velocity v will remain on a straight path if the magnitudes of the electric and magnetic forces are equal. The velocity at which this is achieved can be derived as follows:

$$F_E = F_B \implies QE = QBv \implies v = \frac{E}{B}.$$

By changing the ratio of E to B, charged particles of a certain selected speed may be collected from their undeflected path (hence the name velocity selection).

17 Electromagnetic Induction

17.1 Magnetic Flux and Flux Linkage

Definition 17.1.1. *Magnetic flux* (Φ) is the product of an area and the component of the magnetic flux density perpendicular to that area.

Mathematically,

$$\Phi = B_{\perp}A = BA\cos\theta,$$

where the magnetic flux density B is at an angle θ to the normal of the area A.

The SI unit of magnetic flux is the weber (Wb).

Definition 17.1.2. For a coil of N turns, the *magnetic flux linkage* is defined as the product of the flux Φ and the number of turns N, i.e.

$$\text{magnetic flux linkage} = N\Phi = NBA\cos\theta.$$

17.2 Faraday's Law and Lenz's Law

Electromagnetic induction refers to the phenomenon that a changing magnetic flux could induce an electromotive force in a conductor.

Law 17.2.1 (Faraday's Law)**.** The magnitude of the induced electromotive force in a conductor is proportional to the rate of change of its magnetic flux linkage.

Law 17.2.2 (Lenz's Law)**.** The induced current will flow in a direction that produces effects which oppose the change that induced it.

Faraday's law and Lenz's law can be combined to form the equation

$$E = -\frac{\mathrm{d}(N\Phi)}{\mathrm{d}t}.$$

Note that a current is induced only if there is a complete circuit. An electromotive force is always induced when there is a change in magnetic flux linkage.

Proposition 17.2.3. The magnitude of the electromotive force E induced in a wire of length L moving at velocity v in a uniform magnetic field of flux density B, with both L and v perpendicular to B, is given by

$$|E| = |BLv|.$$

Proof. The area that the wire sweeps out per unit time is given by Lv. Thus,

$$|E| = \left|\frac{\mathrm{d}\Phi}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}(BA)}{\mathrm{d}t}\right| = \left|B\frac{\mathrm{d}A}{\mathrm{d}t}\right| = |BLv|.$$

Note that $N = 1$ since we have only a single conductor (the wire). □

The direction of the induced current in the wire, which is perpendicular to both the magnetic field and the motion of the wire, can be predicted using Fleming's right-hand rule: if the thumb and the first two fingers of the right hand are held so that they are mutually at right-angles, then they represent speed (v), field (B) and current (I) respectively.

17.3 Applications

17.3.1 Generator

Electric generators convert mechanical energy into electrical energy.

When a coil of N turns rotates with a constant angular velocity ω in a uniform magnetic field B, an electromotive force E is induced.

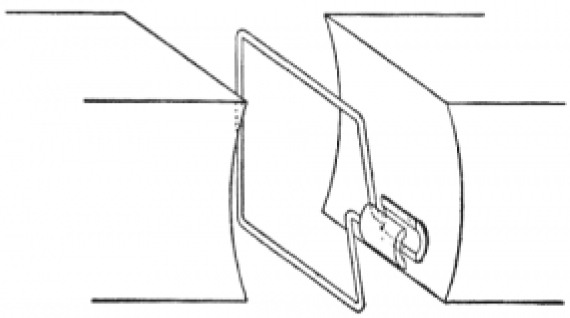

Figure 17.1: An example of a generator.[1]

Assume at time $t = 0$, the coil is perpendicular to the magnetic field. Then $\theta = \omega t$, so the flux linkage through the coil is given by

$$N\Phi = NBA\cos(\omega t)\,.$$

From the laws of induction, the induced electromotive force is given by

$$E = -\frac{\mathrm{d}\,(N\Phi)}{\mathrm{d}t} = NBA\omega\sin(\omega t)\,.$$

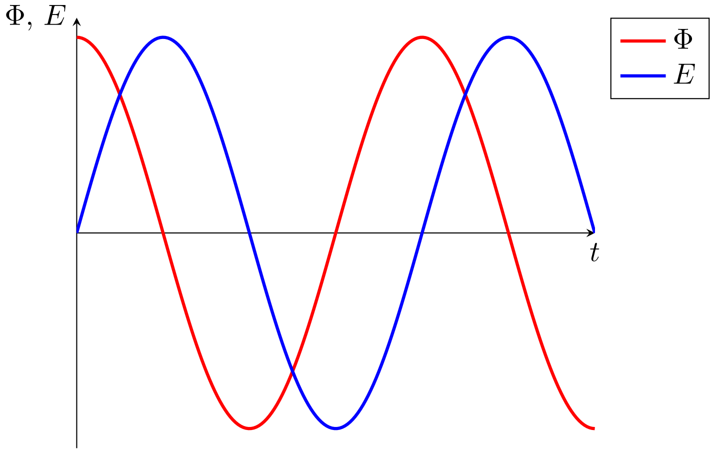

Figure 17.2: A graph of Φ and E over time.

From this equation, we deduce that

- when the flux linking the coil is maximum, the induced electromotive force is zero; and
- when the flux linking the coil is zero, the induced electromotive force is maximum.

[1]Source: https://www.watelectrical.com/electrical-generator-working-types-and-applications/

17.3.2 Eddy Current

According to Faraday's law, any relative motion between a conductor and a magnetic field will give rise to an induced electromotive force in the conductor.

If the magnetic field across a conducting body is not uniform, the induced electromotive force in different parts of the body will be different, giving rise to circulating currents within the body, called *eddy currents*.

According to Lenz's law, the eddy currents will give rise to effects that oppose the relative motion, hence slowing the relative motion. Due to the body's resistance, energy will be dissipated by the eddy currents as thermal energy.

18 Alternating Currents

18.1 Alternating Currents

Definition 18.1.1. An *alternating current* is an electric current that periodically reverses its direction in a circuit.

Definition 18.1.2. The *peak value* (I_0) is the maximum absolute value of the alternating current or voltage in either direction of zero value in a periodic cycle.

Definition 18.1.3. The *peak-to-peak value* is the difference between the positive peak value and the negative peak value of the alternating current within a cycle.

Definition 18.1.4. The *root-mean-square* (I_{rms}) value of an alternating current is the value of the steady direct current that dissipates energy at the same average rate as the alternating current.

Mathematically,

$$I_{\mathrm{rms}} = \sqrt{\frac{1}{T}\int_0^T I^2 \,\mathrm{d}t}.$$

Proposition 18.1.5. The mean power $\langle P\rangle$ dissipated in a load R by an alternating current is related to its root-mean-square current I_{rms} and root-mean-square voltage V_{rms} via

$$\langle P\rangle = I_{\mathrm{rms}}V_{\mathrm{rms}} = I_{\mathrm{rms}}^2 R = \frac{V_{\mathrm{rms}}^2}{R}.$$

Proof. Recall that

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

Replacing I and V by their "mean" values gives us the claim. □

18.1.1 Sinusoidal Alternating Current

The most commonly encountered form of alternating current is the sinusoidal form, i.e.

$$I = I_0 \sin \omega t.$$

Equivalently,

$$V = V_0 \sin \omega t.$$

Proposition 18.1.6. The r.m.s. current I_{rms} and r.m.s. voltage V_{rms} of a sinusoidal alternating current are related to their peak values I_0 and V_0 by

$$\frac{I_{\mathrm{rms}}}{I_0} = \frac{V_{\mathrm{rms}}}{V_0} = \frac{1}{\sqrt{2}}.$$

Proof. Since $I = I_0 \sin \omega t$, we have

$$I_{\mathrm{rms}}^2 = \int_0^T \frac{(I_0 \sin \omega t)^2}{T} \, \mathrm{d}t.$$

Under the substitution $\theta = \omega t$, we obtain

$$I_{\mathrm{rms}}^2 = \frac{I_0^2}{2\pi} \int_0^{2\pi} \sin^2 \theta \, \mathrm{d}\theta = \frac{I_0^2}{2} \implies I_{\mathrm{rms}} = \frac{I_0}{\sqrt{2}}.$$

A similar calculation reveals that

$$V_{\mathrm{rms}} = \frac{V_0}{\sqrt{2}}.$$

□

Proposition 18.1.7. For a sinusoidal alternating current, the mean power $\langle P \rangle$ is half the maximum power P_{max}.

Proof. We have

$$\langle P \rangle = I_{\mathrm{rms}} V_{\mathrm{rms}} = \frac{I_0 V_0}{2} = \frac{P_{\mathrm{max}}}{2}.$$

□

18.2 Transformer

The function of a transformer is to convert one alternating voltage to another of different magnitude.

A simple iron-code transformer comprises a primary and secondary coil of insulated conducting wire wound around a ring of iron.

A changing primary current causes a changing magnetic flux. As a result of electromagnetic induction, a changing electromotive force is produced in the secondary coil.

The coils are wound on an iron core in order to concentrate the magnetic flux and to reduce the flux losses. The iron core ensures that essentially all the magnetic flux is confined to the core and so nearly all the flux passing through the primary coil also pass through the secondary coil. The core is constructed of thin isolated laminations or sheets of soft iron to minimize energy losses due to eddy currents.

An *ideal transformer* is one in which there is no power losses, i.e. the input power is equal to the output power. In addition, the same magnetic flux passes through each turn of both the primary and secondary coils, i.e. there is no flux leakage.

Proposition 18.2.1. In an ideal transformer,

$$\frac{N_S}{N_P} = \frac{V_S}{V_P} = \frac{I_P}{I_S},$$

where N is the number of turns in a coil, and the subscript P and S indicate the primary and secondary coils respectively.

Proof. By the laws of induction, we have

$$V_P = -N_P \frac{\mathrm{d}\Phi_P}{\mathrm{d}t} \quad \text{and} \quad V_S = -N_S \frac{\mathrm{d}\Phi_S}{\mathrm{d}t}.$$

Because the transformer is ideal, we also have $\Phi_P = \Phi_S$, so

$$\frac{V_S}{V_P} = \frac{N_S}{N_P}.$$

Further, the input power is equal to the output power, so

$$I_P V_P = I_S V_S \implies \frac{V_S}{V_P} = \frac{I_P}{I_S}.$$

□

If $N_S > N_P$, the transformer is called a ***step-up*** transformer since $V_S > V_P$. Similarly, if $N_S < N_P$, it is called a ***step-down*** transformer.

Note that a transformer's output and input voltages are not in phase. By Lenz's law, the induced voltage in the secondary coil will produce effects to oppose the alternating flux linkage that induced it, hence it is not in phase with the flux linkage. But the flux linkage in the core varies in phase with the input voltage, so the input and output voltages are not in phase.

18.2.1 Half-Wave Rectification

Rectification is the process of converting an alternating current source into a direct current supply. A half-wave rectification only converts half of the alternating current into direct current by preventing the negative current flow from entering the appliance.

Diodes are used for rectification. The circuit consists of one diode in series with an alternating current input to be rectified and a load requiring direct current input. For simplicity, the load is represented by a resistor R.

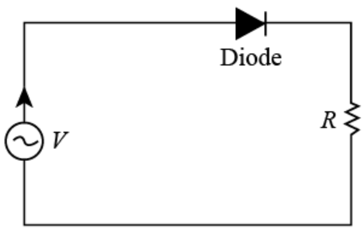

Figure 18.1: A simple half-wave rectifier circuit.[1]

Consider the case of a sinusoidal alternating current. If the first half cycle acts in the forward-biased direction of the diode, current flows round the circuit, creating a potential difference across R which will have almost the same value as the input potential difference V. During the second half cycle, the diode is reverse-biased, hence little or no current flows in the circuit and the potential difference across R is zero. The current is unidirectional and so the potential difference across R is direct, for although it fluctuates, it never changes direction.

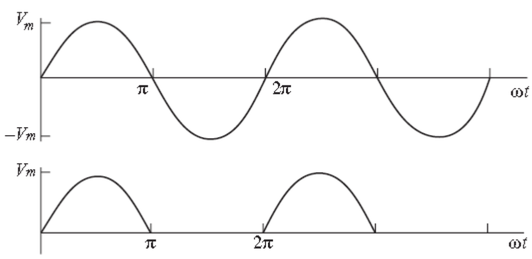

Figure 18.2: The graphs of the input potential difference V and the potential difference across R over time.[2]

[1]Source: https://homework.study.com/explanation/what-is-a-rectifier-circuit.html
[2]Source: https://www.researchgate.net/publication/344782988_Electronic_trainer_for_educational_purposes

Part VI

Modern Physics

19 Quantum Physics

Definition 19.0.1. A *photon* is a discrete packet (or quantum) of energy of electromagnetic radiation.

Proposition 19.0.2. The energy E of a photon of electromagnetic radiation of frequency f is given by

$$E = hf = \frac{hc}{\lambda},$$

where h is the Planck constant.

19.1 Photoelectric Effect

Definition 19.1.1. The *photoelectric effect* refers to the phenomenon that, when certain metal surfaces are illuminated by electromagnetic radiation, electrons are emitted from the surfaces.

The electrons so emitted are called *photoelectrons*.

19.1.1 Einstein's Photon Explanation

If an electromagnetic radiation illuminates a metal surface and ejects an electron, wave theory predicts the following:

- Electrons would be emitted at any frequency, provided the intensity of the radiation is high enough.
- The kinetic energy of the electron should depend on the intensity of the wave.
- Electrons would require some time to absorb incident radiation before they acquire enough kinetic energy to escape from the metal.

However, various observations of the photoelectric effect are not in accordance with the predictions of wave theory. These include:

- No electron is emitted if the frequency of the radiation is below a certain threshold frequency f_0 even with very intense radiation. If the frequency of the radiation is above f_0, electrons are emitted even with low-intensity radiation.
- The maximum energy of the emitted electrons increases with the frequency of the radiation (above f_0) and is independent of the intensity of the radiation.
- There is negligible delay between illumination of the surface and emission of electrons.

Einstein suggested that, in its interaction with matter to release an electron, an electromagnetic radiation behaves as a stream of particle-like photons, each with energy proportional to the frequency of the radiation.

- Each photon delivers a quantum of energy, hf, which could be absorbed by an electron immediately.

- Energy Φ is the minimum energy needed to free an electron from the surface.
  - If hf is less than Φ, no electron is ejected. Higher intensity means more photons per second, but each photon is still unable to eject an electron.
  - If hf is greater than Φ, the remainder is available to the electron as kinetic energy. Lower intensity means fewer photons per second, but each photon is still able to eject an electron.

Definition 19.1.2. The *work function energy* (Φ) of a metal surface is the minimum energy (of a photon) required to release an electron from the surface.

Proposition 19.1.3 (Einstein's Photoelectric Equation)**.** The photoelectric emission can be expressed as

$$hf = \Phi + \frac{1}{2}m_e v_{\max}^2,$$

where m_e is the mass of an electron and $v_{\max}$ is the maximum speed of an emitted electron.

Note that even when the incident light is monochromatic, emitted electrons will have a range of values of kinetic energy up to the maximum value. This is because electrons not at the surface collide with other particles before reaching the surface, thus they require more energy to escape.

Definition 19.1.4. The *threshold frequency* of a metal surface is the minimum frequency an electromagnetic radiation must have in order to cause photoelectric emission.

Proposition 19.1.5. The threshold frequency f_0 is related to the work function Φ by

$$hf_0 = \Phi.$$

Proof. If the radiation has frequency f_0, the emitted photoelectrons will have zero kinetic energy. Thus, $hf_0 = \Phi + 0 = \Phi$. □

Corollary 19.1.6. An equivalent formulation for photoelectric emission is

$$E_k = \frac{1}{2}m_e v_{\max}^2 = hf - hf_0.$$

Proof. From Einstein's photoelectric equation,

$$hf = \Phi - \frac{1}{2}m_e v_{\max}^2 = hf_0 - \frac{1}{2}m_e v_{\max}^2 \implies \frac{1}{2}m_e v_{\max}^2 = hf - hf_0.$$

□

19.1.2 Stopping Potential

Definition 19.1.7. One *electron volt* (eV) is the energy gained by an electron accelerated through a potential difference of one volt.

$$1\ \text{eV} = e \times (1\ \text{V}) = \left(1.60 \times 10^{-19}\ \text{C}\right) \times (1\ \text{V}) = 1.60 \times 10^{-19}\ \text{J}.$$

The electron volt is widely used to express the energy of photons and electrons because its magnitude is convenient in this area.

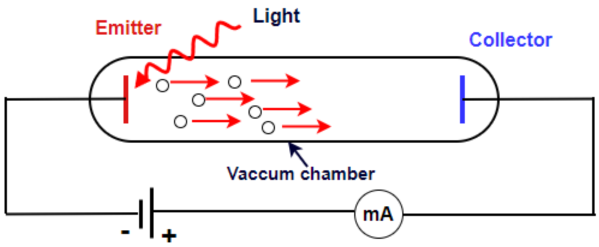

Figure 19.1: Experimental set-up to observe the photoelectric effect.[1]

The above figure shows an experimental set-up in which the photoelectric effect can be observed. An evacuated glass contains a metal electron (emitter, E) connected to the negative terminal of an e.m.f. source. Another metal electrode (collector, C) is connected to the positive terminal. V is the potential of the collector relative to the emitter.

When monochromatic light, of frequency above f_0, shines on plate E, a current is detected by the ammeter, indicating a flow of photoelectrons across the plates E and C.

If the electric field points towards the collector, as shown in the above figure, all the electrons are accelerated toward the collector and contribute to the photocurrent.

However, by reversing the electric field and adjusting its strength as in the figure below, we can determine the maximum kinetic energy of the emitted electrons by making V negative enough so that the current I stops. This occurs for $V = -V_S$, where V_S is called the *stopping potential*.

Proposition 19.1.8. At the stopping potential,

$$eV_S = \frac{1}{2}m_e v_{\text{max}}^2 = hf - h_0.$$

Proof. The most energetic electron leaves the emitter with kinetic energy $\frac{1}{2}m_e v_{\text{max}}^2$ and has zero kinetic energy at the collector. By the conservation of energy, $\Delta\text{KE} + \Delta\text{EPE} = 0$, so

$$\left(0 - \frac{1}{2}m_e v_{\text{max}}^2\right) + (-e)(-V_S) = 0 \implies eV_S = \frac{1}{2}m_e v_{\text{max}}^2.$$

□

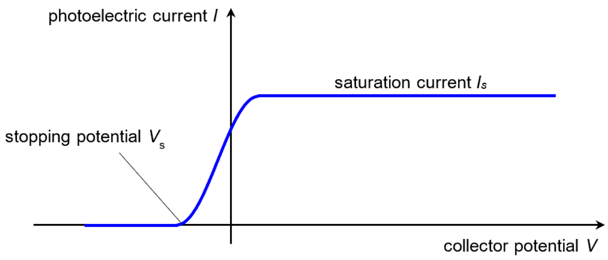

Figure 19.2: A graph of the photoelectric current I against the collector potential V.[2]

As V increases from $-V_S$ towards 0 V, more photoelectrons reach the collector, hence the photocurrent increases. The current-voltage relationship in this region is approximately linear.

At 0 V, there is no accelerating or retarding potential. Hence, only photoelectrons with sufficient initial kinetic energy reach the collector.

[1] Source: https://iplts.com/physics/photoelectric-effect.php
[2] Source: https://xmphysics.com/2023/01/10/17-1-5-i-v-graph/

As V becomes positive, it accelerates all emitted photoelectrons towards the collector. The photocurrent rapidly increases until it reaches a saturation value. Beyond a certain positive voltage, any further increase in voltage will not increase the current.

The saturation current is determined by the intensity of the incident light (number of photons per second), as well as the quantum efficiency of the emitter (electrons emitted per incident photon).

19.2 Wave Particle Duality

Definition 19.2.1. *Wave-particle duality* refers to the fact that matter behaves like waves in some situations and like particles in others.

Definition 19.2.2. The *de Broglie wavelength* (λ) is the wavelength associated with a particle that is moving, given by

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

Definition 19.2.3. The *de Broglie frequency* of a particle is related to its energy E in exactly the same way as for a photon:

$$E = hf.$$

Definition 19.2.4. The momentum p associated with electromagnetic radiation of wavelength λ is given by

$$p = \frac{h}{\lambda} = \frac{hf}{c} = \frac{E}{c}.$$

19.3 Discrete Energy Levels

19.3.1 Energy Levels

Atoms consist of a nucleus surrounded by electrons. In isolated atoms, these electrons occupy specific *energy levels* or *orbitals*, which are characterized by *principal quantum numbers* (n). The principal quantum number is a positive integer that primarily determines the energy of an electron in an atom.

In quantum mechanics, electrons can only exist in certain allowed energy states, which are quantized (they have discrete values). The lowest energy state, corresponding to $n = 1$, is called the *ground state*. Any states above this $n > 1$ are referred to as *excited states*.

The Bohr model provides the simplest illustration of these discrete energy levels. In this model, electrons can only occupy specific circular orbits around the nucleus, each corresponding to a particular energy level and principal quantum number.

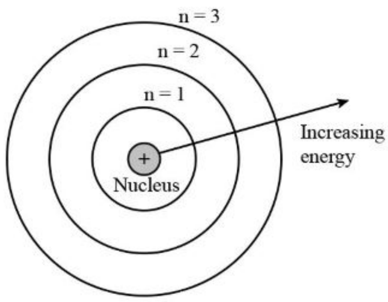

Figure 19.3: The Bohr model of an atom.

19.3.2 Electron Transitions and Spectral Lines

Electrons can move between these energy levels by absorbing or emitting energy. When an electron moves to a higher energy level (increasing n), it absorbs energy in a process called *excitation*. Electrons in excited states are inherently unstable and remain in these higher energy levels for only a very short time. Almost immediately, the electron falls to a lower energy level (decreasing n), releasing energy in a process called *de-excitation*. The energy changes in these transitions occur in the form of photons. The energy of each photon is equal to the energy difference between the levels involved in the transition.

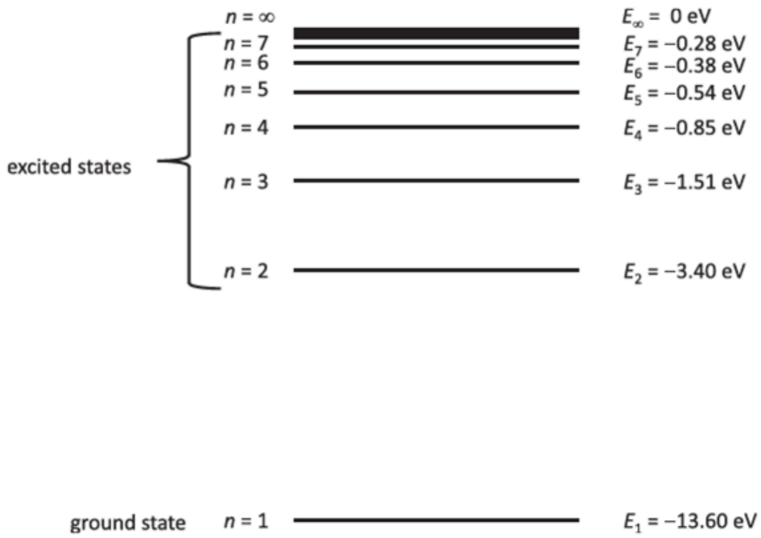

Figure 19.4: The energy level diagram for a hydrogen atom. By convention, the energy of an electron at rest outside the atom is taken as zero.[3]

These electron transitions give rise to *spectral lines*. Each possible transition between energy levels produces a photon of specific energy, corresponding to light of a particular frequency of wavelength. The collection of these specific wavelengths form the atom's *spectrum*. When excited atoms emit photons, we observe an *emission spectrum*, which appears as bright lines on a dark background. Conversely, when atoms absorb specific wavelengths from white light, we see an *absorption spectrum*, appearing as dark lines in a continuous spectrum.

Each element has a unique set of energy levels, resulting in a distinct spectral pattern. This uniqueness allows for identification of elements based on their spectra.

Figure 19.5: The emission and absorption spectrum of hydrogen.[4]

[3]Source: https://ebrary.net/183922/mathematics/energy_levels_transitions
[4]Source: https://montessorimuddle.org/2012/02/01/emission-spectra-how-atoms-emit-and-absorb-light/

19.3.3 X-Ray Spectra

X-rays are electromagnetic radiation with wavelengths ranging from 10^{-12} m to 10^{-9} m. X-rays are ionizing radiation.

Production of X-Rays

X-ray tubes are devices which operate on high voltage in order to generate X-ray photons. A simplified diagram of an X-ray tube is shown below.

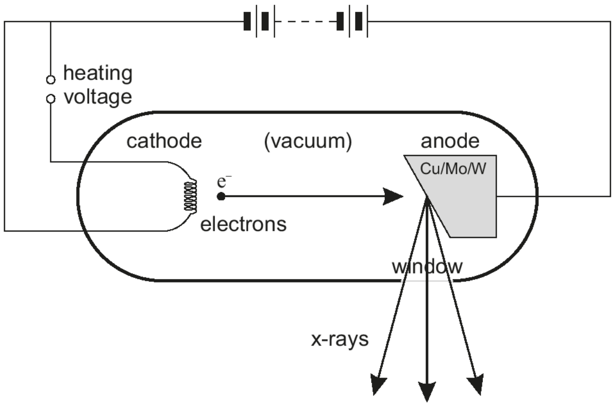

Figure 19.6: A diagram of an X-ray tube.[5]

The filament is heated up by passing electric current through it, and emits electrons. The emitted electron gains kinetic energy given by

$$E = e\Delta V,$$

where ΔV is the potential difference in which the electron is accelerated across. When the electrons that are accelerated by the high voltage collide with a metal target, X-rays are emitted.

X-Ray Spectrum

A typical X-ray spectrum, a graph of intensity I against wavelength λ for the radiation emitted by an X-ray tube, is shown below.

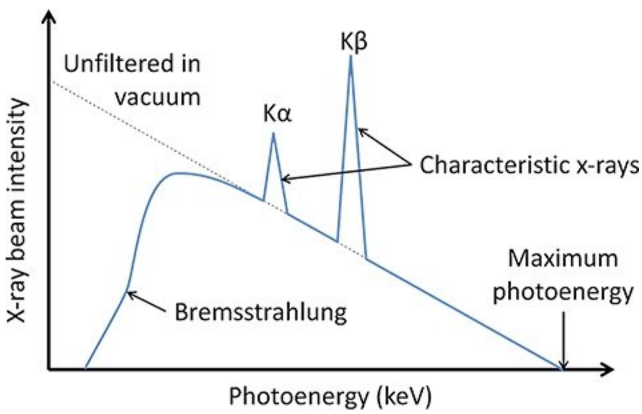

Figure 19.7: An X-ray spectrum.[6]

[5]Source: https://www.oreilly.com/library/view/biomedical-imaging/9783110423518/content/10_chapter03_4.xhtml

It has a continuous broad spectrum called *bremsstrahlung* (lit. braking radiation) and a series of sharp peaks called *characteristic X-rays*.

The bremsstrahlung are emitted as a result of sudden deceleration of the electrons when they collide with the dense metal target. A wide range of decelerations give rise to photons with a wide range of energies and so a continuous spectrum.

The bremsstrahlung cuts off rapidly at a minimum wavelength λ_{min} which is independent of the target metal, but depends on the maximum energy of the electrons, which in turn depends on the accelerating potential difference ΔV. A higher ΔV will produce a shorter cut-off wavelength λ_{min}.

The characteristic X-rays are emitted when electrons transit from higher to lower energy levels in the target atoms. The peaks correspond to the emission spectrum of the target atom. The accelerating potential difference ΔV has to exceed a certain value, usually called the threshold voltage, for the characteristic X-rays to be observed.

19.4 Heisenberg's Uncertainty Principle

Principle 19.4.1 (Heisenberg's Uncertainty Principle)**.** For a particle, the product of the uncertainties in position (Δx) and momentum (Δp) in the same direction cannot be smaller than the Planck constant h:

$$\Delta x \Delta p \geq h.$$

This principle implies that it is impossible to simultaneously determine both the exact position and exact momentum of a particle.

[6] Source: https://physicsopenlab.org/2018/02/13/x-ray-emission/

20 Nuclear Physics

20.1 Nuclear Processes

20.1.1 The Nucleus

Rutherford's experiment consisted of a beam of α-particles (^4_2He), generated by the radioactive decay of radium, directed normally onto a sheet of very thin gold foil in an evacuated chamber.

Alpha particles produce a tiny, but visible flash to light when they strike the fluorescent zinc sulphide screen, which is used as a detector, at the focus of a microscope; the screen and microscope could be swivelled around the foil to observe particles deflected at any given angle.

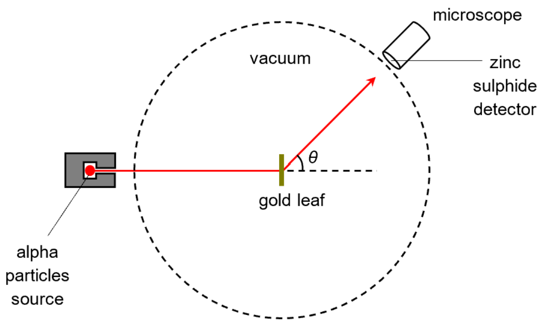

Figure 20.1: The experimental set-up.[1]

The following were observed:

- Most of the α-particles were deviated through small angles. This suggests that the atom is mostly empty space, i.e. the size of the nucleus is small compared to the atom.
- A small but significant percentage of the α-particles were deviated through angles larger than 90°. This suggests that the mass of the atom is concentrated in the nucleus, and that the nucleus is positively charged.

These observations led to the proposal of the nuclear model of an atom, in which positively charged protons and electrically neutral neutrons are tightly packed in a very dense nucleus at the centre. A cloud of negatively charged electrons circulate around the nucleus. The nucleus provides nearly all the mass, but very little of the volume.

Definition 20.1.1. Protons and neutrons in a nucleus are collective known as *nucleons*.

Definition 20.1.2. The *nucleon number* or *mass number* (A) of a nucleus is equal to the number of nucleons in it.

[1] Source: https://xmphysics.com/2023/01/10/18-1-1-rutherford-alpha-particle-scattering-experiment-2/

Definition 20.1.3. The *proton number* or *atomic number* (Z) of a nucleus is equal to the number of protons in it.

Each element in the periodic table has a unique proton number.

All atoms are neutral, hence the proton number of its nucleus and the number of electrons orbiting it must be the same.

Definition 20.1.4. The *unified mass constant* (u) is defined as 1/12 of the mass of a carbon-12 atom.

$$1 \text{ u} = 1.66 \times 10^{27} \text{ kg}.$$

Nucleons have masses of approximately 1 u (proton mass = 1.007 u, neutron mass = 1.008 u), while electrons are much less massive (electron mass = 0.00055 u).

20.1.2 Nuclides and Isotopes

Definition 20.1.5. A *nuclide* is a particular type of nucleus that is specified by its proton number and neutron number.

The usual notation for the representation of a nuclide is

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X},$$

where X is the chemical symbol of the element, A is the nucleon number and Z is the proton number.

Definition 20.1.6. *Isotopes* are atoms/nuclei of the same element that have the same atomic number but different mass numbers.

Isotopes have the same chemical properties but different physical and nuclear properties, e.g. density, magnetic properties and radioactive tendency.

20.1.3 Mass Defect and Binding Energies

Definition 20.1.7. The *mass defect* (Δm) of a nucleus is the difference between the observed mass M of the nucleus and the total mass of the constituent nucleons.

Mathematically,

$$\Delta m = (Zm_p + Nm_n) - M,$$

where m_p is the mass of a proton, m_n is the mass of a neutron, and $N = A - Z$ is the neutron number of the nuclide.

Definition 20.1.8. The *nuclear binding energy* (B, E_b) of a nucleus is the energy required to completely separate the nucleons to infinity. Equivalently, it is the energy released when nucleons come together from infinity.

Proposition 20.1.9. The nuclear binding energy E_b and mass defect of a nucleus Δm is given by

$$E_b = (\Delta m)c^2.$$

The above result is a consequence of the mass-energy equivalence due to Einstein: $E = mc^2$.

20.1.4 Nuclear Reactions

Definition 20.1.10. In a *nuclear reaction*, there is a transformation of at least one element or isotope to another.

A nuclear reaction must be caused by an external stimulus (e.g. the collision of nuclides). Nuclear reactions can be represented by a nuclear equation, such as

$$^{14}_{7}\mathrm{N} + {}^{4}_{2}\mathrm{He} \longrightarrow {}^{17}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{H}.$$

Law 20.1.11 (Conservation Laws in Nuclear Reactions)**.** Nuclear reactions conserve nucleon numbers, proton numbers, mass-energy and linear momentum.

Definition 20.1.12. A nuclear reaction is said to be *exothermic* if there is a net release of energy, and *endothermic* if it requires a net energy input.

The following statements about a nuclear reaction are equivalent:

- The reaction is exothermic (net release of energy).
- The total mass of reactants is more than the total mass of the products.
- The total binding energy of the reactants is less than the total binding energy of the products.

Analogous equivalences are available for endothermic reactions.

20.1.5 Fission and Fusion

Definition 20.1.13. The *binding energy per nucleon* is the average energy per nucleon required to completely separate the nucleons to infinity. Equivalently, it is the average energy released per nucleon when nucleons come together from infinity.

The binding energy per nucleon is a measure of the stability of a nucleus. The higher the binding energy per nucleon, the more stable the nucleus is.

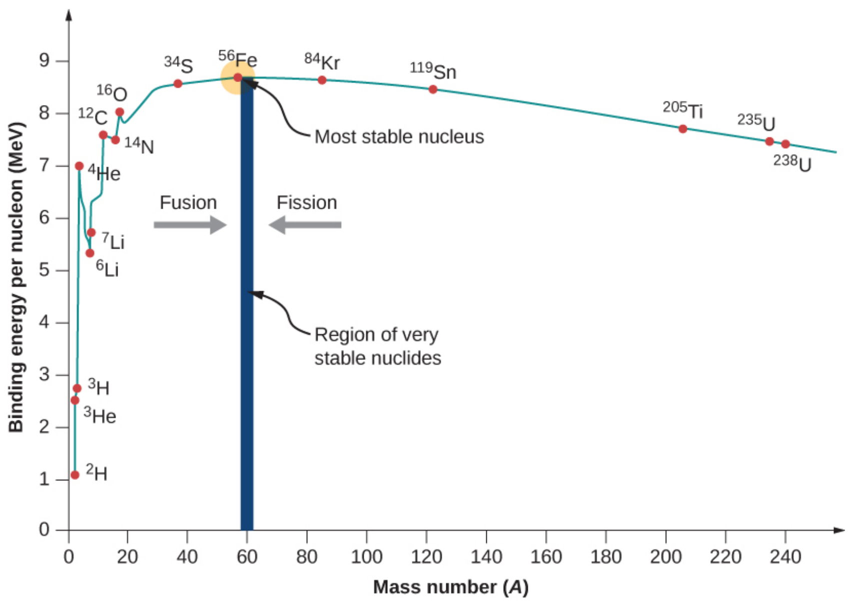

Figure 20.2: The variation of binding energy per nucleon with nucleon number.

The highest binding energy per nucleon (≈ 8.8 MeV) occurs at the region near $A = 56$. Observe further that the slope when $A < 56$ is very steep, while the slope when $A > 56$ is gradual.

For a nuclear reaction to release energy (exothermic), the reactants must have lower binding energy per nucleon than the products. There are two ways this can happen, namely fusion and fission.

Definition 20.1.14. *Nuclear fusion* refers to two light nuclei combining to form a more massive nucleus.

Definition 20.1.15. *Nuclear fission* refers to the breakup of a heavy nucleus into two lighter nuclei with masses of similar orders of magnitude.

Nuclear fission is typically instigated by a neutron. For instance, when a large nucleus like uranium-235 is bombarded by neutrons, one of the many fission reactions that it can undergo is the following:

$$^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \longrightarrow {}^{148}_{57}\mathrm{La} + {}^{85}_{35}\mathrm{Br} + 3\,{}^{1}_{0}\mathrm{n}.$$

Three neutrons are produced in this reaction, which can then go on and hit other nuclei and cause a chain reaction.

Nuclear fission is "easier" to initiate than fusion. In fission, the uncharged neutron can be captured easily by the nuclear due to the strong nuclear force. There is no electrical interaction between the neutron and the positively charged nucleus. However, in fusion, the charged nuclei repel each other, hence they need to have a large kinetic energy before they can get close enough for the strong nuclear force to take effect. The mean kinetic energy of the nuclei is directly proportional to the absolute temperature of the gas, hence high temperatures ($\approx 10^8$ K) are required for the onset of fusion between nuclei.

20.2 Radioactivity

20.2.1 Radioactive Decay

Definition 20.2.1. *Radioactive decay* is the spontaneous and random disintegration of an unstable nucleus into a more stable configuration by emitting α, β and/or γ radiations.

Here, *spontaneous* means that the decay is not affected by external or environmental factors, i.e. it cannot be sped up or slowed down by physical or chemical means. *Random* means that the nucleus has a constant probability of decaying per unit time, hence the time of decay cannot be predicted.

20.2.2 Count Rate

Definition 20.2.2. The *count rate* is a measure of the rate of radiation received by a radioactivity detector.

For a given radioactivity level, the count rates tend to fluctuate irregularly, illustrating the random nature of radioactive decay.

Definition 20.2.3. *Background radiation* is the radiation detected by a radioactivity detector when no radioactive source is nearby.

Background radiation is unavoidable radiation arising from natural resources. It is roughly constant. When measuring the count rate from a source, the background count rate should be deducted from the measured count rate.

20.2.3 α, β and γ Radiation

α Radiation

Definition 20.2.4. *α-particles* contain two protons and two neutrons, i.e. a ${}^{4}_{2}\mathrm{He}$ nucleus.

The typical speed of an α-particle is $0.1c$. It has low penetrating power, and is stopped by a few centimetres of air, or 0.5 mm of paper.

Definition 20.2.5. *α-decay* is the spontaneous decay of a nucleus with the emission of an α-particle.

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \longrightarrow {}^{\mathrm{A}-4}_{\mathrm{Z}-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He}.$$

Since X is unstable compared to Y, the mass of the reactant is more than the mass of the products and the decay is exothermic.

Proposition 20.2.6. In α-decay

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \longrightarrow {}^{\mathrm{A}-4}_{\mathrm{Z}-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He},$$

let K_α and m_α be the kinetic energy and mass of the α-particle respectively, and K_Y and m_Y the kinetic energy and mass of the daughter nucleus respectively. Then

$$\frac{K_\alpha}{K_\alpha + K_Y} = \frac{m_Y}{m_\alpha + m_Y}.$$

Proof. By the conservation of linear momentum,

$$m_\alpha v_\alpha = m_Y v_Y \implies \frac{v_\alpha}{v_Y} = \frac{m_Y}{m_\alpha},$$

where v_α and v_Y are the speeds of the a-particle and daughter nucleus respectively. Then

$$\frac{K_\alpha}{K_Y} = \frac{\frac{1}{2}m_\alpha v_\alpha^2}{\frac{1}{2}m_Y v_Y^2} = \frac{v_\alpha}{v_Y} = \frac{m_Y}{m_\alpha}.$$

Thus,

$$\frac{K_\alpha}{K_\alpha + K_Y} = \frac{\frac{K_\alpha}{K_Y}}{\frac{K_\alpha}{K_Y} + 1} = \frac{\frac{m_Y}{m_\alpha}}{\frac{m_Y}{m_\alpha} + 1} = \frac{m_Y}{m_\alpha + m_Y}.$$

$\square$

From the above result, we see that in α-decay, most of the released energy $(K_\alpha + K_Y)$ is carried by the α-particle in the form of kinetic energy.

β Radiation

Definition 20.2.7. *β-particles* are high-speed electrons emitted from within the nucleus of an atom.

The beta particles reach speeds of up to $0.9c$. It is stopped by a few metres of air, or a few millimetres of aluminimm.

Definition 20.2.8. *β-decay* is the spontaneous decay of a nucleus with the emission of a β-particle.

During β-decay, a neutron changes into a proton and an electron. The new proton is retained in the nucleus, while the electron is ejected as a β-particle.

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \longrightarrow {}_{\mathrm{Z}+1}^{\mathrm{A}}\mathrm{Y} + {}^{0}_{-1}\mathrm{e} + \overline{\nu}.$$

Here, $\overline{\nu}$ represents an *antineutrino*. It has zero electric charge, and its mass is a tiny fraction of the mass of the electron.

In β-decay, the total energy Q released is shared, in varying proportions, between the emitted electron and the neutrino.

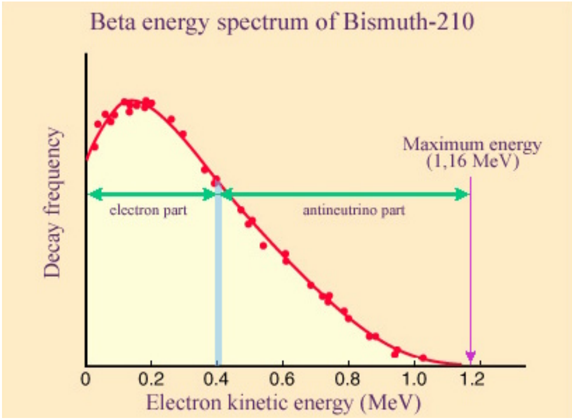

Figure 20.3: The β-particle energy distribution for bismuth-210.[2]

γ Radiation

Definition 20.2.9. *γ-radiation* is short-wavelength electromagnetic radiation.

γ-radiation has strong penetrating power. It is stopped by kilometre of air, or 10 cm of lead.

Definition 20.2.10. *γ-decay* is the spontaneous decay of a nucleus with the emission of gamma radiation.

Frequently, some of the energy released in α- and β-decay is in the form of a γ-ray photon. This is because α- and β-decay usually leave the daughter nucleus in an excited state (denoted by *). The daughter nucleus emits energy in the form of a γ-ray photon in order to reach its ground state.

$${}^{A}_{Z}\mathrm{X}^* \longrightarrow {}^{A}_{Z}\mathrm{X} + \gamma.$$

20.2.4 Activity

Definition 20.2.11. The *activity* (A) of a radioactive sample is the rate at which nuclei decay, i.e. the number of disintegrations per unit time.

Mathematically,

$$A = -\frac{\mathrm{d}N}{\mathrm{d}t},$$

where N is the number of undecayed particles.

The SI unit of activity is the Becquerel (Bq).

Definition 20.2.12. The *decay constant* (λ) is the probability of decay per unit time.

The SI unit of the decay constant is second^{-1} (s^{-1}).

Activity is directly proportional to the number of undecayed particles N. The decay constant is the constant of proportionality between A and N:

$$A = \lambda N.$$

[2]Source: https://radioactivity.eu.com/articles/phenomenon/beta_spectrum

Proposition 20.2.13. Given a sample of initial activity A_0, its activity A after time t is given by $A = A_0\mathrm{e}^{-\lambda t}$.

Proof. By definition,

$$A = \lambda N = -\frac{\mathrm{d}N}{\mathrm{d}t},$$

which is a first order differential equation with solution $N = N_0 e^{-\lambda t}$, where N_0 is the original number of undecayed radioactive nuclei in the sample. Since $A = \lambda N$, we multiply through by λ to get $A = A_0\mathrm{e}^{-\lambda t}$. □

Any detector placed at a certain fixed position to measure the activity of a radioactive sample will receive a certain fraction of the radiations emitted from that sample. Thus, the received count rate C (= measured count rate − background count rate) is directly proportional to activity A:

$$C \propto A.$$

Definition 20.2.14. The ***half-life*** $(t_{1/2})$ of a radioactive nuclide is the average time taken for its activity to fall to half its initial value.

Proposition 20.2.15. The half-life $t_{1/2}$ is related to the decay constant λ by

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

Proof. At $t = t_{1/2}$, the activity $A = A_0/2$. Thus,

$$\frac{1}{2}A_0 = A_0\mathrm{e}^{-\lambda t_{1/2}} \implies \mathrm{e}^{\lambda t_{1/2}} = 2 \implies \lambda t_{1/2} = \ln 2 \implies t_{1/2} = \frac{\ln 2}{\lambda}.$$

□

Corollary 20.2.16. The activity A after time t can be written as

$$A = A_0\left(\frac{1}{2}\right)^{t/t_{1/2}}.$$

Proof. We have

$$A = A_0\mathrm{e}^{-\left(\frac{\ln 2}{t_{1/2}}\right)t} = A_0\left(\mathrm{e}^{-\ln 2}\right)^{t/t_{1/2}} = A_0\left(\frac{1}{2}\right)^{t/t_{1/2}}.$$

□