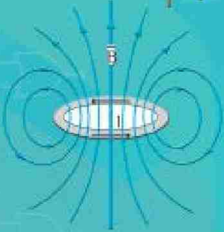
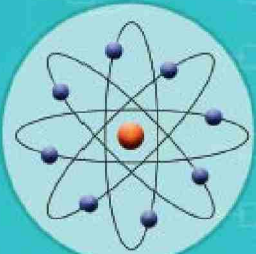
Physics

for Advanced Level Secondary Schools

Student's Book

Form Six





$$V_0 = \left(\frac{R_1+R_2}{R_1}\right)\left(\frac{R_4}{R_4+R_3}\right)V_2 - \frac{R_2}{R_1}V_1$$

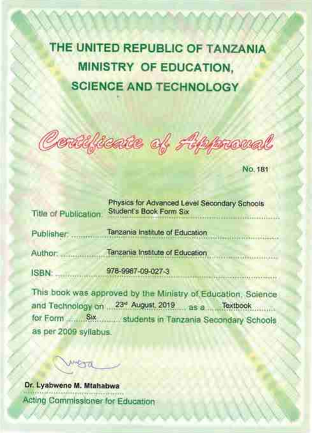
Tanzania Institute of Education

Physics

for Advanced Level Secondary Schools

Student's Book

Form Six



THE UNITED REPUBLIC OF TANZANIA
MINISTRY OF EDUCATION,
SCIENCE AND TECHNOLOGY

Certificate of Approval

No. 181

Title of Publication: Physics for Advanced Level Secondary Schools Student's Book Form Six

Publisher: Tanzania Institute of Education

Author: Tanzania Institute of Education

ISBN: 978-9987-09-027-3

This book was approved by the Ministry of Education, Science and Technology on 23rd August, 2019 as a Textbook for Form Six students in Tanzania Secondary Schools as per 2009 syllabus.

Dr. Lyabwene M. Mtahabwa
Acting Commissioner for Education

Tanzania Institute of Education

Published 2019

ISBN 978 – 9987 – 09 – 027 – 3

Tanzania Institute of Education
P. O. Box 35094
Dar es Salaam

Tel: 255 22 2773005/+255-22-2771358
Fax: +255 22 2774420
Email: director.general@tie.go.tz
Website: www.tie.go.tz

Preface

This book, ***Physics for Advanced Level Secondary Schools,*** is written specifically for Form Six students in the United Republic of Tanzania. The book is prepared according to the 2009 Physics Syllabus for Advanced Secondary Education Form V - VI issued by the Ministry of Education and Vocational Training.

The book consists of five chapters, which are: Current electricity, Electromagnetism, Electronics, Modern physics and Environmental physics. In addition, to the content, each chapter contains illustrations, exercises, activities, worked examples, and revision questions. Answers to numerical questions are provided at the end of the book. Learners are encouraged to do all practical work and answer all questions. This will enhance their understanding, and promote the acquisition of the intended skills and competencies for Form six students.

Tanzania Institute of Education

Acknowledgements

The Tanzania Institute of Education (TIE) would like to acknowledge the contribution of all organizations, and individuals who participated in the design and development of this textbook.

Writers: Prof. Peter Msaki, Dr Yohana Msambwa, Dr Mwingereza Kumwenda, Dr Godson Lema, Dr Frank Tilya, Mr Malima Nyamwaga, Mr Godbless Shao & Mr Ramadhani Mntambo

Editors: Dr Christian Uiso & Dr Nuru Mlyuka (Chairperson of the panel)

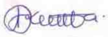
Designer: Mr Amani Kweka

Illustrators: Mr Halifa Halifa & Alama Art and Media Production Company Ltd.

Coordinator: Dr Godson Lema

TIE also extends its sincere gratitude to the secondary school teachers who participated in the trial phase of the manuscript.

Likewise, the Institute would like to thank the Ministry of Education, Science and Technology for facilitating the process of writing this textbook.



Dr Aneth A. Komba
Director General
Tanzania Institute of Education

Table of contents

Preface iii
Acknowledgement iv

Chapter One : Current electricity 1
1.1 Basic concepts 1
1.2 Electric conduction in metals 2
1.3 Electric conduction in gases 21
1.4 Alternating current theory 28

Chapter Two : Electromagnetism 58
2.1 Magnetic fields and forces 58
2.2 Sources of magnetic fields 73
2.3 Electromagnetic induction 81
2.4 Magnetic properties of materials 92
2.5 Magnetic field of the earth 100

Chapter Three : Electronics 109
3.1 The band theory of solids 109
3.2 Classification of semiconductors 114
3.3 Transistors 125
3.4 Logic gates 142
3.5 Operational amplifiers 149
3.6 Telecommunication 160

Chapter Four : Modern physics 174
4.1 Quantum physics 174
4.2 Atomic physics 187
4.3 Laser 195
4.4 Nuclear physics 198

Chapter Five : Environmental physics 215
5.1 Agricultural physics 215
5.2 Energy from the environment 228
5.3 Earthquakes 242
5.4 Environmental pollution 249

Answers 259
Table of physical constants 263
Glossary 265
Bibliography 271
Index 273

Chapter One

Current electricity

Introduction

Life would be very difficult without the use of appliances and devices that are powered by current electricity. The study of current electricity is the first step towards understanding how these appliances and devices work. In order to design or repair the electrical devices used in daily life it is important to be acquainted with the knowledge of electrical circuit components such as resistors, capacitors and inductors. In this chapter, you will learn about the origins of current electricity, how is conducted in different electric circuits, and how the devices used for measuring electricity across components are constructed and used.

1.1 Basic concepts

Current electricity or electric current is due to the flow of electric charges. The SI unit of electric current is the ampere (A). The ampere is a fundamental unit which gives a measure of the rate of flow of electrons in an electrical conductor.

The magnitude of a current in a circuit is equal to the rate of flow of charge (Q) through a conductor. For a steady current (I):

$$I = \frac{Q}{t} \qquad (1.1)$$

where t is time in seconds (s) and Q is the charge in Coulombs (C).

When the current is not steady, the instantaneous current (I) is given by the following formula:

$$I = \frac{dQ}{dt} \qquad (1.2)$$

Since the charge is discrete, then $Q = Ne$, where N is the number of electrons flowing in the conductor, and e is the charge of an electron. Then the steady current can be expressed as:

$$I = \frac{Ne}{t} \qquad (1.3)$$

and instanteneous current, as

$$I = \frac{d}{dt}(Ne) \qquad (1.4)$$

1.2 Electric conduction in metals

Electric conduction in metals is carried by free electrons. Metals have a structure such that each atom has one or more electron(s) occupying the valence shell. Free electrons have thermal energy which depends on the metal temperature and move randomly throughout the structure with an average speed of approximately 10^6ms^{-1}.

When a power supply is connected across the ends of a metal conductor an electric field $(\vec{E})$ is developed in the region occupied by the metal. The free electrons (Ne) are affected by this field and are forced to move in the opposite direction towards the point of higher potential (positive terminal of the power supply). Hence, the applied potential difference across the conductor induces a small drift velocity (about 10^{-3}ms^{-1}) on the motions of the free electrons. The flow of electric charges across the cross-section area A of the conductor is the one that produces an electric current (I), shown in Figure 1.1.

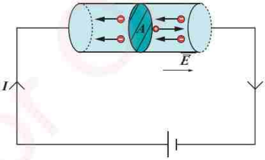


Figure 1.1 Motion of the electrons in the metallic conductor

1.2.1 Drift velocity of electrons

Drift velocity, v_d is an average velocity with which the free electrons move towards the positive terminal under the effects of applied electric field. Refer to a portion of a conductor, for example copper wire, through which the current (I) is flowing in Figure 1.2. Let n be the number of free electrons per unit volume of the conductor; n can be expressed as follows:

$$n = \frac{N}{Al}$$

$$N = nAl \tag{1.5}$$

where N is the total number of electrons in a volume Al with A and l representing the cross-sectional area and length of the conductor respectively.

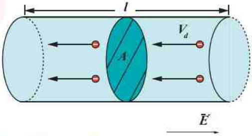


Figure 1.2 Illustration of electric conduction in metal

To obtain current through this conductor we substitute equation (1.5) into (1.3) to get:

$$I = \frac{Ne}{t} = \left(\frac{nAl}{t}\right)e \quad \text{since} \quad v_d = \frac{l}{t} \quad \text{then}$$

$$I = nev_d A$$

Therefore, $$v_d = \frac{I}{neA} \tag{1.6}$$

Example 1.1

Suppose, $I = 5\,A$, Area = 2 mm^2
$e = 1.6 \times 10^{-19} C$ and
$n = 10^{28}$ electrons / m^3. Obtain the drift velocity of the electron.

Solution

The drift velocity of electrons can be obtained as:

$$v_d = \frac{I}{neA}$$

$$= \frac{5\text{A}}{10^{28}\,\text{electrons/m}^3 \times 1.6 \times 10^{-19}\text{C} \times 2 \times 10^{-6}\,\text{m}^2}$$

$$v_d = 1.5625 \times 10^{-3}\,\text{ms}^{-1}$$

Thus, the drift velocity is very small compared with the average thermal speeds of the free electrons in a conductor.

Example 1.2

Find the total momentum (P) acquired by the electrons in a wire of length 35 cm when a current of 2.5 A starts to flow.

Solution

$$P = mv_d$$

but $M = Nm_e$

$$P = Nm_e \times \frac{IAl}{NeA} = \frac{m_e Il}{e}$$

$$P = \frac{9.1 \times 10^{-31}\text{kg} \times 2.5\text{A} \times 0.35\text{m}}{1.6 \times 10^{-19}\text{C}}$$

$$= 4.98 \times 10^{-12}\,\text{kgms}^{-1}$$

Thus, the total momentum acquired by the electrons is 4.98×10^{-12} kgms^{-1}

1.2.2 Current density (*J*)

Consider a wire of uniform cross-sectional area (A) carrying a current (*I*). The current density (J) is defined as the electric current per unit cross-sectional area of a conductor given by:

$$J = \frac{I}{A} \qquad (1.7)$$

Since $I = nev_d A$ then J can be expressed as:

$$J = nev_d \qquad (1.8)$$

The potential difference (*V*) across the conductor produces electric field (*E*) which exerts electric force $F_e = eE$ on the free electrons. According to Newton's 2nd law this force is given by:

$$F_e = -eE = ma$$

$$\text{a} = -\frac{eE}{m}$$

The negative sign indicates that direction of acceleration is opposite to that of electric field.

The drift velocity (v_d) acquired by an electron accelerated from rest during relaxation time (τ) is given by:

$$v_d = -\left(\frac{eE}{m}\right)\tau \qquad (1.9)$$

Using $E = -\frac{V}{l}$, and equations (1.6) in (1.9), and rearranging terms, the current through the conductor is given by:

$$I = \left(\frac{ne^2 A}{ml}\tau\right)V \qquad (1.10)$$

If A, τ and l are constant, then

$$V = \left(\frac{ml}{ne^2 A\tau}\right)I = IR \qquad (1.11)$$

Equation (1.11) is commonly known as Ohm's law, where R is a proportionality constant or resistance of a conductor with SI unit of Ohm (Ω).

The reciprocal of resistance denoted by letter G is called the conductance.

$$G = \frac{1}{R} \qquad (1.12)$$

The SI unit of conductance is per Ohm (Ω^{-1}). Conductors which obey Ohm's law are called ohmic or linear conductors Figure 1.3 (a), for example pure metals, copper sulphate solution etc, and those which do not obey Ohm's law are called non-ohmic conductors Figur 1.3 (b), for example junction diodes, vacuum diodes etc.

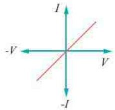


Figure 1.3(a) I-V characteristics of ohmic conductors.

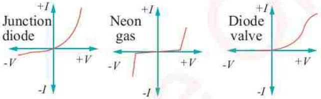


Figure 1.3(b) I-V characteristics of non-ohmic conductors.

From equations (1.10) and (1.11), R can be expressed as:

$$R = \frac{ml}{ne^2 A\tau} \qquad (1.13)$$

1.2.3 Resistivity of a conductor

From equation (1.13), it is important to note that the resistance of a conductor is directly proportional to the length and inversely proportional to the cross-sectional area (A) of a conductor, that is:

$$R \,\alpha\, \frac{l}{A}$$

The constant of proportionality ρ is called resistivity of the material. Thus:

$$R = \rho\frac{l}{A} \qquad (1.14)$$

Since both equations (1.13) and (1.14) represent resistance, it is evident that resistivity of the conductor is given by:

$$\rho = \frac{m}{ne^2\tau} \qquad (1.15)$$

Since the relaxation time (τ), is material and temperature dependent these quantities are expected to have influence on resistance according to equation 1.15. Resistivity of some selected materials at a room temperature is presented in Table 1.1.

Table 1.1 Resistivity of selected materials at room temperature

Materials	Resistivity (Ωm)
Silver	1.6×10^{-8}
Copper	1.7×10^{-8}
Aluminium	2.8×10^{-8}
Iron	9.8×10^{-8}
Constantan	4.9×10^{-7}
Mercury	9.8×10^{-7}
Germanium	4.6×10^{-1}
Tin	1.09×10^{-7}
Silicon	6.4×10^{2}
Fused quartz	7.5×10^{17}
Brass	8.0×10^{-8}
Tungsten	5.6×10^{-8}
Carbon (graphite)	$(2.5\text{-}5) \times 10^{-6}$
Nichrome	$1.1 \text{ x } 10^{-6}$

1.2.4 Conductivity

Conductivity of a conductor denoted by σ is defined as the reciprocal of the resistivity given by:

$$\sigma = \frac{1}{\rho} = \frac{l}{AR} \qquad (1.16)$$

The SI unit of conductivity is $\Omega^{-1}\text{m}^{-1}$.

Insert equation (1.16) in equation (1.11) and rearranging terms, the following expression can be obtained:

$$I = \frac{A\sigma}{l}V$$

$$\frac{I}{A} = \frac{V}{l}\sigma$$

It is important to note that $\frac{I}{A} = J$ and $\frac{V}{l} = E$ as defined earlier, hence

$$J = E\sigma \qquad (1.17)$$

Since, electrical conductivity is analogous to thermal conductivity it implies that a good electrical conductor is also a good heat conductor.

1.2.5 Temperature dependence of resistance

It is known that the resistance of a conductor is a function of temperature. Suppose the resistance of a conductor at temperature θ_o is R_o and the temperature increases to θ when resistance changes to R_θ. Therefore, the change in the resistance $\left(\Delta R = R_\theta - R_o\right)$ is directly proportional to R_θ, directly proportional to the temperature change $\left(\Delta\theta = \theta - \theta_o\right)$ and depends on the nature of the materials.

From the above explanations

$$\Delta R \propto R_o \Delta\theta$$

$$\Delta R = \alpha R_o \Delta\theta$$

where α is proportionality constant called temperature coefficient of resistance; it is defined as the fractional change in the resistance of a conductor per degree rise in temperature. Thus:

$$R_\theta - R_o = \alpha R_o\left(\theta - \theta_o\right)$$

Hence:

$$R_\theta = R_o\left[1+\alpha\left(\theta-\theta_0\right)\right] \quad (1.18)$$

where R_θ and R_o are both measured in Ohm and θ in °C. If θ_o is assumed to be zero i.e. $\theta_o = 0°C$, then

$$R_\theta = R_o\left(1+\alpha\theta\right) \quad (1.19)$$

A graph of the resistance, R versus temperature, θ will result into Figure 1.4 with the resistance intercept as the original resistance, R_o and gradient is the product of R_o and the temperature coefficient of the resistance, α.

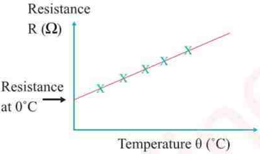


Figure 1.4 Variation of resistance R against temperature θ of a conductor.

Suppose R_1 and R_2 are resistances at temperature θ_1 and θ_2 respectively, then substituting them in equation (1.19) gives;

$$R_1 = R_o\left(1+\alpha\theta_1\right) \quad (1.20)$$

$$R_2 = R_o\left(1+\alpha\theta_2\right) \quad (1.21)$$

Taking the ratio of equations (1.20) and (1.21) we obtain;

$$\frac{R_1}{R_2} = \frac{\left(1+\alpha\theta_1\right)}{\left(1+\alpha\theta_2\right)} \quad (1.22)$$

If R_1, R_2, θ_1 and θ_2 are known then α can be determined.

Example 1.3

The windings of a motor has a resistance of 80 Ω at 15 °C. Find its resistance at 50 °C given that the temperature coefficient of the winding material is 0.004 / °C.

Solution

From equation (1.22)

$$\frac{R_{15}}{R_{50}} = \frac{\left(1+\alpha\theta_1\right)}{\left(1+\alpha\theta_2\right)}$$

$$\frac{80}{R_{50}} = \frac{\left(1+0.004\times15\right)}{\left(1+0.004\times50\right)}$$

$$R_{50} = 90.57\ \Omega$$

Therefore, the resistance at 50 °C is 90.57 Ω.

Example 1.4

A copper wire has a resistance of 1.81 Ω at 20 °C and 2.24 Ω at 80 °C. Find:

(a) the temperature coefficient of a copper; and

(b) the resistance of the wire at 60 °C.

Solution

From

$$R_{20} = R_o\left(1+20\alpha\right) \quad (i)$$

$$R_{80} = R_o\left(1+80\alpha\right) \quad (ii)$$

Dividing equation (i) by (ii) we get;

$$\frac{R_{20}}{R_{80}} = \frac{R_o(1+20\alpha)}{R_o(1+80\alpha)}$$

$$\alpha = \frac{R_{80} - R_{20}}{80R_{20} - 20R_{80}}$$

$$= \frac{(2.24 - 1.81)\ \Omega}{(80\ ^{\circ}C \times 1.181\ \Omega) - (20\ ^{\circ}C \times 2.24\ \Omega)}$$

$$= 4.3 \times 10^{-3}\ \text{K}^{-1}$$

Therefore, the temperature coefficient of copper is $4.3 \times 10^{-3}\ \text{K}^{-1}$

(b) From $\frac{R_{20}}{R_{60}} = \frac{R_o(1+20\alpha)}{R_o(1+60\alpha)}$

$$R_{60} = \left(\frac{1+(4.3\times10^{-3}\ \text{K}^{-1}\times 60\ ^{0}C)}{1+(4.3\times10^{-3}\ \text{K}^{-1}\times 20\ ^{0}C)}\right)$$

$$\times 1.81\ \Omega = 2.10\ \Omega$$

Thus, the resistance of a wire at 60 °C is $2.10\ \Omega$.

1.2.6 Electrical networks

An electric circuit needs a source of power to produce current. Such devices are sources of electromotive force (e.m.f.) denoted by E. The e.m.f. of a battery is equal to the potential difference across the source terminals when the circuit is open. A battery gives a nearly constant voltage; however, it has a small internal resistance (r), which reduces the actual voltage from the ideal (e.m.f.). The terminal voltage of the battery, V is the voltage that is measured across the terminal of the battery Figure 1.5. This is defined as the difference between the e.m.f. of the battery, E and the potential drop across the internal resistance of the battery, that is

$$V = E - Ir \qquad (1.23)$$

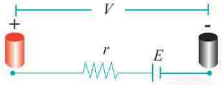


Figure 1.5 Terminal voltage of a d.c. circuit

The e.m.f. of the battery is due to the nature of the chemicals used and not on the size of a battery. The small battery has the same e.m.f. as that of a large battery made from the same chemicals. The internal resistance of the small battery is however much less than that of the large battery, provided only a small current is taken from a battery, its internal resistance and e.m.f. are almost constant.

(a) Arrangement of resistors and cells

Resistors can be connected in parallel or in series depending on the requirements. For instance parallel and series connection can be observed in household wirings and industrial installations.

(i) Series combination

In this combination, two or more resistors are connected consecutively as illustrated in Figure 1.6.

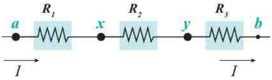


Figure 1.6 A series combination of three resistors

Since these resistors are in series, the same current I, flows in all three resistors, so that:

$I_1 = I_2 = I_3 = I$ where I_1, I_2 and I_3 are currents through R_1, R_2 and R_3 respectively.

Also the sum of the potential differences across the three resistors is equal to the potential difference between point a and b; that is:

$$V_{ab} = V_{ax} + V_{xy} + V_{yb}$$

From Ohm's law

$$V_{ax} = IR_1; \quad V_{xy} = IR_2;$$

$$V_{yb} = IR_3 \text{ and } V_{ab} = IR_{eq}$$

Since the current I is common then it cancels to get:

$$R_{eq} = R_1 + R_2 + R_3$$

This equation gives us the relation for calculating the equivalent resistance of the three resistors connected in series. The equivalent resistance (R_{eq}) for a sequence of n resistors in series is just the sum of the individual resistances.

Generally the equivalent resistance is the sum of resistors given as:

$$R_{eq} = \sum_{i}^{n} R_i \qquad (1.24)$$

(ii) Parallel combination

In parallel combination, each terminal of a resistor is connected across the battery, as illustrated in Figure 1.7. It can be noted that the potential difference across each resistor has to be the same.

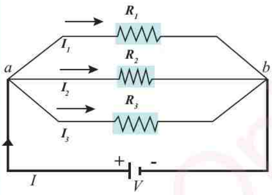


Figure 1.7 A parallel combination of three resistors

Note also that, the current I is equal to the sum of the currents I_1, I_2 and I_3 so that:

$$I = I_1 + I_2 + I_3 \text{ and } I_1 = \frac{V_{ab}}{R_1}, \quad I_2 = \frac{V_{ab}}{R_2},$$

$$I_3 = \frac{V_{ab}}{R_3} \text{ and } I = \frac{V_{ab}}{R_{eq}}$$

Substitution of current and observing the potential is common then, equivalent resistance is given by:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

As before the reciprocal equivalent resistance is given by the sum:

$$\frac{1}{R_{eq}} = \sum_{i}^{n} \frac{1}{R_i} \qquad (1.25)$$

Example 1.5

In the circuit shown in Figure 1.8, evaluate the following:

(i) equivalent resistance of the circuit, and;

(ii) current flowing through each resistor.

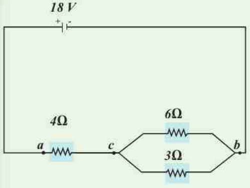


Figure 1.8 An electric circuit

Solution

(i) Since the two resistors between terminals c and b are in parallel then the equivalent resistance is:

$$\frac{1}{R_{eq}} = \frac{1}{6\,\Omega} + \frac{1}{3\,\Omega} = \frac{1}{2\,\Omega}$$

$$R_{eq} = 2\,\Omega$$

Since the equivalent resistor is in series with the 4 Ω resistor they can be combined as follows $R_{eq} = 4\,\Omega + 2\,\Omega = 6\,\Omega$

Therefore, the equivalent resistance for the circuit is 6 Ω.

(ii) The current flowing through the 4Ω resistor is calculated using the equivalent resistance and the potential drop across it using Ohm's law

$$I_{4\Omega} = \frac{V}{R_{eq}} = \frac{18\,\text{V}}{6\,\Omega} = 3\,\text{A}$$

The 3 Ω and 6 Ω resistors are in parallel, then the $p.d.$, (V_{cb}) across each is the same.

Hence

$$E = V_{ac} + V_{cb}$$

$$V_{cb} = E - V_{ac} = 18\,\text{V} - (4\,\Omega \times 3\,\text{A})$$

$$V_{cb} = 6\,\text{V}$$

Therefore currents through $3\,\Omega$ and $6\,\Omega$ resistors are;

$$I_{3\Omega} = \frac{V}{R_3} = \frac{6\,\text{V}}{3\,\Omega} = 2\,\text{A} \quad \text{and}$$

$$I_{6\Omega} = \frac{V}{R_6} = \frac{6\,\text{V}}{6\,\Omega} = 1\,\text{A}$$

(b) Combination of cells

Like resistors, batteries can also be connected in parallel or in series.

(i) Series connection

In this arrangement, the positive terminal of the first cell is connected to the negative terminal of the second cell and so on as illustrated in Figure 1.9

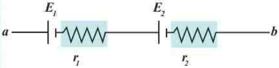


Figure 1.9 Combination of cells in series

The equivalent e.m.f. is the sum of the individual e.m.f. of the respective cells.

$$E_{eq} = E_1 + E_2$$

while the equivalent internal resistance, r_{eq} is the sum of the individual internal resistance of the respective cells.

$$r_{eq} = r_1 + r_2$$

In general, the equivalent e.m.f. for n cells arranged in series is given by the sum:

$$E_{eq} = \sum_{i=1}^{n} E_i \qquad (1.26)$$

The equivalent internal resistances for n cells arranged in series is given by:

$$r_{eq} = \sum_{i=1}^{n} r_i \qquad (1.27)$$

(ii) Parallel connection of the cells

In this arrangement, like poles are connected as shown in Figure 1.10

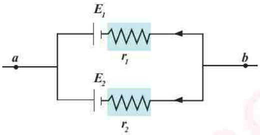


Figure 1.10 Combination of cells in parallel

For cells with equal e.m.f. the effective e.m.f. is considered as the e.m.f. of one of the cells.

Thus, $E_{eq} = E_1 = E_2$

while the equivalent internal resistance, r_{eq} can be calculated using equation (1.25).

Hence; $\dfrac{1}{r_{eq}} = \dfrac{1}{r_1} + \dfrac{1}{r_2}$

For n cells with similar e.m.f. arranged in parallel is equal to the e.m.f. of an individual cell that is

$$E_{eq} = E_1 = E_2 = E_{n-1} = E_n \qquad (1.28)$$

The equivalent internal resistance, r_{eq} for n cells with equal e.m.f. arranged in parallel is given by the sum:

$$\frac{1}{r_{eq}} = \sum_{i=1}^{n} \frac{1}{r_i} \qquad (1.29)$$

The equivalent *e.m.f.* and internal resistance for non-equivalent n cells arranged in parallel are calculated using Kirchhoff's laws which will be described in the following section.

1.2.7 Kirchhoff's laws

Potential differences and steady currents in series and parallel networks are usually complicated systems of electrical circuits. Kirchhoff extended the ideas of Ohm's laws and came out with two laws which together enable to solve the electrical network problems.

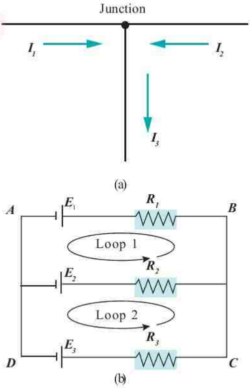


Figure 1.11 Junction and loops

A junction (also called a node or branch point) is a point where three or more electric conductors meet as illustrated in Figure 1.11 (a) while a *loop* is any closed conducting path as depicted in Figure 1.11(b).

(a) Kirchhoff's current law (KCL) is sometimes known as the **junction theorem which** states that; "The algebraic sum of the currents at any junction is zero". This statement is expressed as the sum:

$$\sum_{i=1}^{n} I_i = 0 \qquad (1.30)$$

The law implies that there is no accumulation of electric charges at any junction. Thus Kirchhoff's current law implies the conservation of electric charges at the junction.
It is important to note the sign convention used in this law, which states that; "A current flowing towards a junction is considered to be positive and vice versa also is true". For the junction given in Figure 1.11 (a)

$$I_1 + I_2 - I_3 = 0$$

(b) Kirchhoff's voltage law (KVL) states that; "In any closed path or loop in the network the algebraic sum of the e.m.f. (E) is equal to algebraic sum of the product of the current and the resistance". The law can be expressed as the sum

$$\sum E = \sum IR \qquad (1.31)$$

It is necessary that one direction should be chosen in traversing any loop that is either clockwise or anticlockwise direction. The e.m.f. is positive if we pass from the positive terminal of the power supply through the external resistor back to the negative terminal and vice versa. The potential differences are positive when there is a drop of a potential and negative when there is a potential rise. Kirchhoff's voltage law is based on the law of conservation of electrical energy.
Kirchhoff's laws can be illustrated as shown in Figure 1.12.

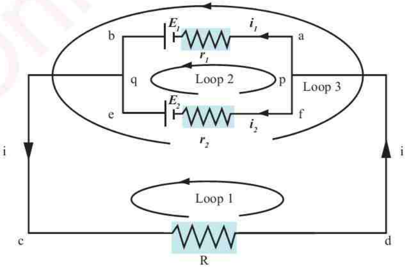


Figure 1.12 Combination of cells and external resistor

By applying KCL, the node equation for loop 2 can be written as:

$$i = i_1 + i_2$$

Also by applying the KVL the following equations can be obtained;

Lets consider loop 1 (cdpfeqc)

$$E_2 = iR + i_2 r_2$$

For loop 2 (abqefpa)

$$E_1 - E_2 = i_1 r_1 - i_2 r_2$$

For Loop 3 (abqcdpa)

$$E_1 = i_1 r_1 + iR$$

When these laws (KCL and KVL) are used they generate a set of simultaneous equations that can be solved using standard procedures.

Example 1.6

Use Figure 1.13, write the corresponding equations that can allow you to obtain the three currents I_1, I_2 and I_3.

Solution

Loop I, (afeba) we have;

$$-I_1R_1 - I_1R_2 - I_3R_4 + E_1 - E_3 = 0 \qquad (i)$$

Loop II, (bedcb) we have;

$$-I_2R_3 + I_3R_4 - E_2 + E_3 = 0 \qquad (ii)$$

Loop III, (afedcba) we have

$$-I_1R_1 - I_2R_3 - I_1R_2 - E_2 + E_1 = 0 \qquad (iii)$$

Junction equation at b gives;

$$I_1 - I_2 - I_3 = 0 \qquad (iv)$$

Assume that the batteries are:

$E_1 = 19\,V$; $E_2 = 6V$; and $E_3 = 2\,V$

and the resistors are:

$R_1 = 6\ \Omega$; $R_2 = 4\ \Omega$; $R_3 = 4\ \Omega$; $R_4 = 1\ \Omega$.

$$-6I_1 - 4I_1 - I_3 = -19 + 2 = -17$$

$$-4I_2 + I_3 = 6 - 2 = 4$$

$$-6I_1 - 4I_2 - 4I_1 = 6 - 19 = -13$$

$$I_1 = I_2 + I_3 \qquad (v)$$

$$10I_2 + 11I_3 = 17 \qquad (vi)$$

$$-4I_2 + I_3 = 4 \qquad (vii)$$

By solving equations (v) to (vii) the values of the currents are $I_1 = 1.5$ A; $I_2 = -0.5$ A; and $I_3 = 2.0$ A. The negative sign on I_2 indicates that the current is actually in the opposite direction to that shown on the diagram.

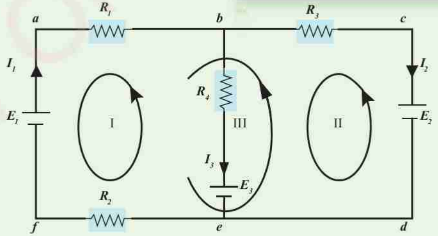


Figure 1.13 Electric circuits

1.2.8 The Wheatstone bridge

Wheatstone bridge circuit is a method of measuring small resistance compared to that measured by ammeter-voltmeter technique. It involves making adjustments until a galvanometer is undeflected. This being a null method, it does not depend on the accuracy of instruments. A Wheatstone bridge is an arrangement of four resistors namely R_1, R_2, R_3 and R_4 connected to form a bridge circuit with four junctions A, B, C and D as shown in Figure 1.14. A power supply is connected across the bridge circuit through points A and B, and the sensitive galvanometer is connected between the other two points C and D. The current I from the power supply branches at junction A into I_1 and I_3 which pass through R_1 and R_3, respectively. At junction C, current I_2 branches to I_2 and I_g which pass through R_2 and the sensitive galvanometer respectively. At junction D, the current I_3 combines with I_g resulting to I_4 which passes through R_4.

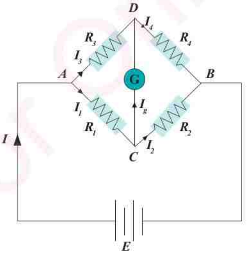


Figure 1.14 Wheatstone bridge

On adjusting one of the resistors until there is no deflection (null deflection) in the galvanometer, the Wheatstone bridge is then said to be *balanced i.e.* the current through the galvanometer is zero ($I_g = 0$). In this case, the current in R_2 is the same as that in R_1 (i.e. $I_1 = I_2$) and the current in R_3 is the same as that in R_4 (i.e. $I_3 = I_4$). As there is no current in the galvanometer, the potential difference across the terminals C and D is zero.

$$V_C = V_D$$

Applying Ohm's law to R_1 and R_2,

$$V_{AC} = I_1R_1 \quad \text{and} \quad V_{CB} = I_1R_2$$

$$\therefore \quad \frac{V_{AC}}{V_{CB}} = \frac{R_1}{R_2}$$

Now applying Ohm's law to R_3 and R_4, and using the same procedure as in the previous relation, we get:

$$V_{AD} = I_3R_3 \quad \text{and} \quad V_{DB} = I_3R_4$$

$$\therefore \quad \frac{V_{AD}}{V_{DB}} = \frac{R_3}{R_4}$$

As the potential at points C and D are equal, then these equations yields:

$$\frac{V_{AC}}{V_{CB}} = \frac{V_{AD}}{V_{DB}} \quad \text{and} \quad \frac{R_1}{R_2} = \frac{R_3}{R_4}$$

The last equation is the condition only satisfied when the Wheatstone bridge is balanced. If one of the resistor is unknown say R_1 and the other remaining resistors R_2, R_3 and R_4 are known then R_1 can be evaluated from the relation:

$$R_1 = \frac{R_3}{R_4}R_2 \qquad (1.32)$$

1.2.9 Wheatstone meter bridge

The Wheatstone meter bridge consists of a wire running on a scale of length 100 cm. This wire with uniform cross sectional area stretched tight between two metallic strips, as shown in Figure 1.15. The metallic strips have two gaps where known and unknown resistors can be connected. The end points (A and B) where the wire is clamped are connected to a cell through a switch. One end of a galvanometer is connected to the metallic strip midway (point C) between the two gaps. The other end of the galvanometer is connected across terminals C and D via a metallic knife-edge device (jockey) which can slide over the wire to make electrical connection.

The unknown resistor, R_1 whose resistance is to be found is connected across one of the gaps, a standard known resistance R_2 is connected across the other gap. The jockey touching point D on the wire, a distance l_1 (cm) from the end A can be slided along the wire in search of a null point D. The portion AD of the wire has a resistance R_{AD}, using equation (1.14)

$R = \rho \dfrac{l}{A}$, it follows that $R_{AD} \propto l_1$

Also the portion DB $(100 - l_1)$ of the wire similarly has a resistance which implies that:

$R_{DB} \propto (100 - l_1)$

The four arms AD, DB, AC and CB obviously form a Wheatstone bridge with AB as the battery arm and CD the galvanometer arm. If the jockey is moved along the wire, then there will be one position where the galvanometer will show no current. Let the distance of the jockey from the end A at the balance point be l.

In this case, the current in R_2 is the same as that in R_1 (i.e. $I_1 = I_2$) and the current along AD of resistance R_{AD} is the same as that along DB of resistance R_{DB} (i.e. $I_3 = I_4$). As there is no current in the galvanometer, the potential difference across the terminals is zero, i.e.

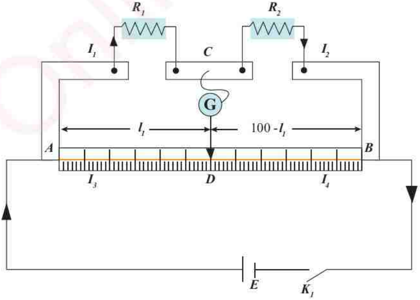


Figure 1.15 The Wheatstone meter bridge

$$V_C = V_D$$

Applying Ohm's law to R_1 and R_2,

$$V_{AC} = I_1 R_1 \quad \text{and} \quad V_{CB} = I_1 R_2$$

$$\therefore \quad \frac{V_{AC}}{V_{CB}} = \frac{R_1}{R_2}$$

Now applying Ohm's law to R_{AD} and R_{DB}, and using the same procedure as in the previous relation, one gets:

$$V_{AD} = I_3 R_{AD} \quad \text{and} \quad V_{DB} = I_3 R_{DB}$$

$$\therefore \quad \frac{V_{AD}}{V_{DB}} = \frac{R_{AD}}{R_{DB}} = \frac{l}{100 - l}$$

Since the potential at points C and D are equal, the ratios of voltages and resistances are equal and can be given by:

$$\frac{V_{AC}}{V_{CB}} = \frac{V_{AD}}{V_{DB}}$$

$$\frac{R_1}{R_2} = \frac{l}{100 - l}$$

If one of the resistors is unknown, say R_1, and the resistor R_2 the balance length l are known, then R_1 can be calculated from the relation:

$$R_1 = \left(\frac{l}{100 - l}\right) R_2 \qquad (1.33)$$

Example 1.7

In a meter bridge Figure 1.15, the null point is found at a distance of 40 cm from A. If a resistance of 15 Ω is connected in parallel with R_2 and null point occurs at 53 cm, determine the values of R_1 and R_2.

Solution

Based on Figure 1.15 the relation between resistances and lengths is given as:

$$\frac{R_1}{R_2} = \frac{l}{100 - l}$$

From the first balance point, we obtain;

$$\frac{R_1}{R_2} = \frac{40\,\text{cm}}{60\,\text{cm}}$$

If a 15Ω resistor is connected in parallel with R_2, the equivalent resistance is

$$R_{2eq} = \frac{15\,\Omega \times R_2\,\Omega}{15\Omega + R_2\,\Omega}$$

From the new balance point, it follows:

$$\frac{R_1}{R_{2eq}} = \frac{53\text{ cm}}{47\text{ cm}} = \frac{R_1\,\Omega \times (15\,\Omega + R_2\,\Omega)}{15\,\Omega \times R_2\,\Omega}$$

$$\frac{53\text{ cm}}{47\text{ cm}} = \frac{R_1\,\Omega}{R_2\,\Omega} \times \frac{(15\,\Omega + R_2\,\Omega)}{15\,\Omega}$$

$$= \frac{40\,\text{cm}}{60\,\text{cm}}\left(\frac{(15\Omega + \text{R}_2\,\Omega)}{15\Omega}\right)$$

$$R_2 = \left(\frac{53\text{ cm} \times 60\text{ cm}}{47\,\text{cm} \times 40\text{ cm}} \times 15\,\Omega\right) - 15\,\Omega$$

$$= 10.37\,\Omega$$

The value of R_1, is obtained by using any of the above expression:

$$R_1 = \frac{40\text{ cm}}{60\text{ cm}} \times R_2\,\Omega = \frac{40\text{ cm}}{60\text{ cm}} \times 10.37\,\Omega$$

$$= 6.91\,\Omega$$

Therefore, the values of the resistances are $\text{R}_1 = 6.91\,\Omega$ and $\text{R}_2 = 10.37\,\Omega$.

1.2.10 Potentiometer

A potentiometer is an accurate device for measuring e.m.f. of the cell or the potential difference between two points of an electric conductor. In its simplest form it consists of resistance wire AB of uniform cross-sectional area lying alongside a ruler with millimeter scale. Through the wire a steady current is maintained by a driving cell as illustrated in Figure 1.16. Also the potentiometer can be used for the determination of the internal resistance of the cell and comparison of the *e.m.f.s* of different cells.

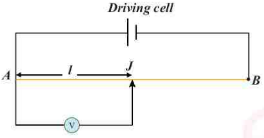


Figure 1.16 A potentiometer

A potentiometer can be adopted to measure current and resistance of a wire. In principle if a wire of uniform cross-section area A is carrying a steady current (I), a fall of the potential difference across any portion of the wire is directly proportion to the length (l) of that portion. Based on Figure 1.16, if the potential difference across AJ is V and the length of AJ is l then, the resistance across AJ is given by:

$$R_{AJ} = \frac{\rho l}{A}$$

From Ohm's law the *p.d.* across AJ becomes:

$$V = IR = \frac{I\rho l}{A} = \left(\frac{I\rho}{A}\right) l$$

Since I, A, and ρ are constants then:

$$V \propto l \qquad (1.34)$$

Equation (1.34), which indicates that the potential drop across a portion of the conductor is proportional to the length of that portion, becomes the principle behind the potentiometer operation.

(a) Comparison of the e.m.f. of cells

In some cases, it becomes necessary to compare e.m.f. of a cell using a device whose circuit diagram is shown in Figure 1.17. Point d represents a sliding contact that is used to vary the resistance (and hence the potential difference) between points a and d. Other required components are the galvanometer, the battery of known e.m.f. (E_o), and the battery of unknown e.m.f. (E_x). Based on the Kirchhoff's junction law we observe that, I is the current in the left hand branch and I_x is the current in the right hand branch.

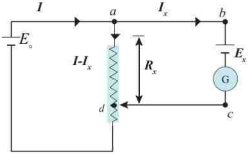


Figure 1.17 The circuit diagram of potentiometer

Applying Kirchhoff's loop law to the loop "abcda" traversed clockwise gives:

$$-E_x + (I - I_x)R_x = 0$$

Because current I_x passes through it, the galvanometer displays a nonzero reading. The sliding contact at d is now adjusted until the galvanometer reads zero indicating that the circuit showing in Figure 1.17 is another null-measurement device. Under this condition, the current in the galvanometer is zero, and the potential difference between *a* and *d* must be equal to the unknown e.m.f. (E_x):

$$E_x = \left(I - I_x\right) R_x$$

Next, the battery of unknown e.m.f. is replaced by a standard battery of known *e.m.f.* (E_s), and the procedure is repeated. If R_s is the resistance between *a* and *d*, and when the balance is achieved we have:

$$E_s = \left(I - I_s\right) R_s$$

In this case I_s is the current through the galvanometer.

At balanced condition (null deflection), I_x and I_s are zero, hence I remains unchanged.

If we write

$$E_x = IR_x$$

and

$$E_s = IR_s$$

The equation enable us to get the unknown e.m.f.

$$E_x = \frac{R_x}{R_s} E_s \qquad (1.35)$$

If the resistor in Figure 1.17 is a wire of resistivity ρ, its resistance can be varied by using the sliding contact to vary the length l. Then, the principle of potentiometer, $E_x \propto l_x$ and $E_s \propto l_s$ equation (1.35), becomes:

$$E_x = \frac{l_x}{l_s} E_s \qquad (1.36)$$

where l_x is the resistor length corresponding to use of the battery of unknown *e.m.f.*, (E_x) in the circuit and l_s is the resistor length related to the use of the standard battery of e.m.f., E_s.

The sliding-wire circuit of Figure 1.17 without the unknown *e.m.f.* and the galvanometer is also called a voltage divider. This circuit makes it possible to tap into any desired smaller portion of the *e.m.f.*, E_0 by adjusting the length of the resistor.

Example 1.8

A standard cell of e.m.f. 1.50 V gives a balance length of 52 cm when connected across the potentiometer. If a dry cell of unknown e.m.f. (E) replaces the standard cell the new balance length of 75 cm on the wire is obtained. Find the value of E.

Solution

From equation (1.36);

$$E_x = \frac{l_x}{l_s} E_s$$

Substituting the given data

$$E_x = \frac{75\,\text{cm}}{52\,\text{cm}} \times 1.5\ \text{V} = 2.16\ \text{V}$$

Hence, the unknown e.m.f. is 2.16 V.

(b) Measurement of the internal resistance of a cell

The potentiometer is also used to measure internal resistance of a cell by first finding the balance length l (AP) for cell when a switch K is open Figure 1.18. Thus, the *p.d.* across AP equals to the *p.d.* between the terminals of the cell, and therefore E_x is proportional to length l. A known resistance R is then connected across the cell by closing switch K and if l_1 (AQ) is the new balance length, the *p.d.*, V which maintains current through R is proportional to l_1:

$$\frac{E_x}{V} = \frac{l}{l_1}$$

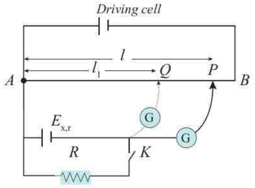


Figure 1.18 Circuit diagram of potentiometer

If the current through R at balance is I and r is the internal resistance of the cell, Ohms law applied first to the whole circuit and then to R alone gives:

$E_x = I(R+r)$ and $V = IR$

$$\frac{E_x}{V} = \frac{R+r}{R} = \frac{l}{l_1}$$

$$r = \left(\frac{l-l_1}{l_1}\right)R \qquad (1.37)$$

Hence the internal resistance r can be calculated if l, l_1 and R are known.

Example 1.9

In measuring the internal resistance of a cell the e.m.f. is first balanced by a length 90 cm on the wire carrying a constant current. When a resistance of 5 Ω is connected across the cell, the terminal potential different of the cell is now balanced by length 45 cm.

a) Why is the balance length smaller when a resistor is connected across the cell?

b) Calculate the internal resistance of the cell.

Solution

a) Some *p.d.* is used in driving the current through the internal resistance, thus some of the *p.d.* appears across the internal resistance.

b) From equation (1.37)

$$r = \left(\frac{l-l_1}{l_1}\right)R = \left(\frac{90\,\text{cm} - 45\,\text{cm}}{45\,\text{cm}}\right)5\Omega = 5\Omega$$

Therefore the internal resistance of the cell is 5 Ω

(c) Energy transfer in an electric circuit

In time t, the charge $q = It$ flows through the circuit. As this charge moves from one point to another, the electric potential energy decreases by:

$$U = qV = It(IR) = I^2Rt$$

This loss in electric potential energy appears as an increase in thermal energy of the resistor. Thus, the current I for a time t through a resistor R increases the thermal energy by I^2Rt. The rate at which a device converts electrical potential energy to other forms is known as electrical power of a device, which is calculated using equation (1.38).

$$P = \frac{U}{t} = I^2 R = VI = \frac{V^2}{R} \qquad (1.38)$$

Activity 1.1

Given a meter bridge, connecting wires, galvanometer, and jockey, one standard resistor of known resistance, one unknown resistor battery and key. Determine the value of the unknown resistor by connecting an appropriate circuit, and using the procedures we have learned for this task.

Activity 1.2

You are provided with a potentiometer, a galvanometer, jockey, resistance box, key, two dry cells in series and one a dry cell whose internal resistance r is to be determined. Carry out experimental procedures to find the value of r.

Exercise 1.1

1. Consider positive and negative charges moving horizontally through four regions shown in Figure 1.19.

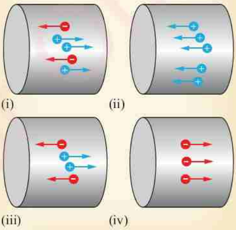


Figure 1.19 Movements of charges

Discuss the mechanism behind the current conduction through metallic conductors and rank the current in these four regions from lowest to highest.

2. What does it mean by the term current density? Establish the relationship between current density J, and electric intensity E. A negligibly small current is passed through a wire of length 15 m and uniform cross-sectional area $6 \times 10^{-7}\,\text{m}^2$ and its resistance is measured to be 5 Ω. What is the resistivity of the material?

3. An overhead cable consists of a copper core enclosed by six straight parallel steel wires each of the same diameter as the copper core. If a current of 16 A flows through this cable, find the current which flows through the steel and the copper. The resistivities of steel and copper are $2.1 \times 10^{-7}\,\Omega\,\text{m}$ and $5.0 \times 10^{-9}\,\Omega\,\text{m}$ respectively.

4. Define the term drift velocity. What are the factors that determine the drift velocity of charge carriers? Estimate the current density and average drift velocity of conduction electrons in copper wire of cross-sectional area $2 \times 10^{-7}\,\text{m}^2$ carrying a current of 0.5 A. (Assume that each copper atom contributes roughly one conduction electron. Density of Copper $= 9 \times 10^3\,\text{kg/m}^3$ and its atomic mass $= 63.5$ g / mole).

5. Power transmission lines operate at very high voltages whereas household circuits operate at fairly low voltages. Justify this statement. Two wires of equal length, one of aluminium and the other of copper have the same resistance. Which of the two wires is lighter? (Given that, resistivity $\rho_{Al} = 2.63 \times 10^{-18}\,\Omega\text{m}$ $\rho_{Cu} = 1.72 \times 10^{-8}\,\Omega\text{m}$ and the relative density of $Al = 2.7$ and that of $Cu = 8.9$).

6. Find the value of the effective resistance and current I in the following circuit shown in Figure 1.20.

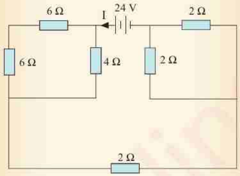


Figure 1.20 An electrical network

7. Given that each resistor in Figure 1.21 has a resistance of 1 Ω, find:
   a) the equivalent resistance; and
   b) the current delivered by a 10 V supply having an internal resistance of 0.8 Ω.

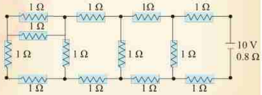


Figure 1.21 An electrical network

8. Consider an experiment carried out at 0 °C on two wires, X which is a Nichrome wire of resistivity $1 \times 10^{-6}\,\Omega\text{m}$ and diametre 1.20 mm and Y which is 1.60 m of a Germanium alloy 0.80 mm in diametre and resistivity $2.8 \times 10^{-7}\,\Omega\text{m}$. The ratio of resistances of wire X to that of wire Y is 1.5. What was the length of the Nichrome wire? If the temperature coeficient of Nichrome is $4 \times 10^{-4}\,K^{-1}$ and that of Germanium alloy is $3 \times 10^{-4}\,\text{K}^{-1}$. What would be the ratio of resistances if the temperature was raised by 80 K?

9. In an experiment to investigate the variation of resistance with temperature, a nickel wire and a 10 Ω standard resistor were connected in the gaps of the Wheatstone metre bridge. When the nickel wire was at 0 °C a balance point was found 40 cm from the end of the bridge wire adjacent to the nickel wire. When it was at 100 °C, the balance point was at 50 cm.

   a) Calculate:
      (i) the temperature of the nickel wire (on its resistance scale) when the balance point was at 42 cm from the end of the bridge wire adjacent to the nickel wire; and
      (ii) the resistivity of nickel at this temperature if the wire was 150 cm long and of cross-sectional area of 2.5×10^{-4} cm^2.

   b) Explain the advantage of using a 10 Ω standard resistor in preference to 100 Ω standard resistor in this experiment.

10. Use Figure 1.22 below, calculate:
    (a) the current through R_1, R_2 and R_3.
    (b) the *p.d.* between *a* and *b*.

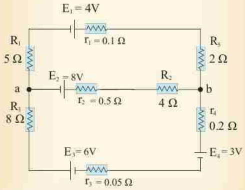


Figure 1.22 An electrical network

1.3 Electric conduction in gases

The conduction of electricity in gases is due to electrical discharge of the gas whereby ions and electrons are generated through different mechanisms. Gases consist of neutral molecules under normal conditions. In order for the gas to pass current, it must first get ionized and hence create charges and an electric field should exist to give direction to the motion of the electric charges. For any gas at a given temperature and pressure there is a minimum value of voltage applied, known as a *breakdown potential* at which ionization takes place and the gas conducts electricity.

Devices such as sodium vapour lamps, florescent lights and processes such as arc welding utilize electric conduction in gasses. In this sub section some of the mechanisms through which electric conduction in gasses takes place, the optical spectrum of gasses as well as applications of conduction of electricity in gases will be discussed.

1.3.1 Conduction of electricity in gases

The conductivity of a gas increases as its pressure is reduced; to produce sparks at a certain normal pressure, a very large potential difference is required. However, at low pressure a small potential difference will create current. To study the electric discharge in air at low pressure a high potential difference is applied across two electrodes at opposite ends of partially evacuated glass tube as displayed in Figure. 1.23.

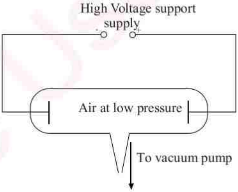


Figure 1.23 An electric discharged tube

The appearances of the discharge at various stages are shown in Figure 1.24 and can be described as follows:

(a) Violet streamers pass between cathode and anode.
(b) A pink positive column and a negative glow appear near the cathode.
(c) The positive column becomes striated, the negative glow moves away from the cathode.
(d) Crookes' dark space extended to fill the whole tube.

Immediately after the pump is switched on the pressure is reduced in the tube. At about 20 mmHg a brush discharge at the electrodes is observed and violet streamers are observed linking the two electrodes. Broadening of the streamers into a deep salmon pink discharge is observed when the pressure is further reduced, this practically fills the space between the electrodes as shown in Figure 1.24 (a).

When the pressure is about 5 mmHg, a dark region appears near the cathode and the column is separated into two parts, the pink positive column ending in the anode glow, and a blue negative glow near the cathode. This dark region is known as *Faraday's dark space* as appeared in Figure 1.24 (b).

Further evacuation to about 0.1 mmHg results into positive column shrinking towards the anode and starts to break into striations. The Faraday dark space and the negative glow increases in length, thereafter a second dark region appears near the cathode, this is known as *Crookes' dark space*. The rest of the tube is filled with pink positive column. It should be clear from Figure 1.24 (c) that the length of the dark spaces do not depend on the length of the tube, but depends only on the pressure.

When the pressure is about 0.01 mmHg the pink positive column and the negative glow have disappeared and the Crookes dark space extend to fill the whole of the tube, during this stage the walls of the tube show a green fluorescence as depicted in Figure. 1.24 (d). In this type of discharge the colour obtained depends on the nature of the gas and materials of the glass.

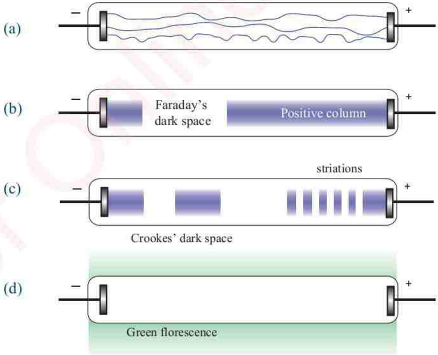


Figure 1.24 Appearance of the discharge at various stages

Ionized gases

At a standard temperature and pressure gases are electrical insulators and about 30 kV is required to produce a spark discharge between two rounded electrodes 1cm apart in air. For pointed electrodes 1 cm apart the value is about 12 kV.

For a gas to conduct electricity it must be ionized. Ionization results into two types of charged particles, i.e. positive ions and negative electrons. On applying the potential difference between the two electrodes in an ionized gas, positive ions move towards the cathode and electrons towards the anode thus, both ions and electrons being charged, a current flows in the gas.

The ionizing agents in gases are categorized as follows:

Nuclear: Nuclear transformations of radioactive materials e.g. radium, results into emission of alpha (α) and beta (β) particles as well as gamma (γ) rays, which can ionize gases, because they usually have high energy.

X-rays and Ultraviolet light: Photon of x-rays have energy between 10^2-10^6 eV while those of ultraviolet light is between 1-10^2 eV. Both types can ionize gases.

Accelerated particles: Proton or an electron or any singly charged ion that has been accelerated through a potential difference can cause ionization.

The ionization processes

There are two types of process as an atom can be ionized namely, photon absorption and collision with charged particles. When a gas is exposed to ultra-violet radiation of sufficient short wave-length, X-rays or the γ -radiation from radioactive bodies, ionization by photon occurs. When charged particles have sufficient kinetic energy, they ionize atoms by collision with atomic electrons.

At ordinary pressure, when it is required to produce ionization in a sample of gas, the most convenient agents are UV radiation, X-rays, or γ -radiation. For instance, the latter, is used in the textile industry to prevent the build-up of electrostatic charges on fabrics, since the charges leak away rapidly when the neighbouring air is ionized.

The three possibilities for gas ionization to occur are:

- A gas is subjected to steady continuous ionization throughout its volume at ordinary pressure. To some extent this approaches the conditions found in electrolytes at low potential gradients, the current is carried by ions and migrate all over the body of the gas. On average the distance travelled between collisions by an ion is small, so that the effect of ionization by collisions will be insignificant unless the potential gradient is large.
- For a gas enclosed in a discharge tube at reduced pressure, as the pressure is lowered, the average distance

travelled by an accelerated ion is longer; and ionization by collision becomes significant even at fairly low potential gradients.

- When the pressure in a discharge tube is extremely low, close to a high vacuum, few gas molecules that supply ions generated at or near the electrodes will travel across the tube without collision. Therefore at this pressure, there are few gas molecules to furnish ions, and these must be provided at one or both of the electrodes.

Ionization curve

Air between anode and cathode electrodes in a discharged tube is ionized by a beam of ionizing radiation as shown in Figure 1.25 (a). A sensitive current device (G) (example a *d.c.* amplifier or a pulse electroscope) is used to record the ionization current.

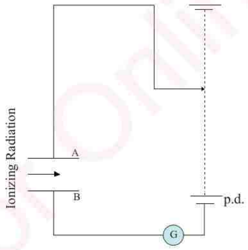


Figure 1.25 (a) Discharged tube

The potential difference can be varied by adjusting the potential divider resulting into ionization current while the intensity of radiation remain constant. A plot of ionization current against potential difference results into curve OPQRS, Figure. 1.25 (b).

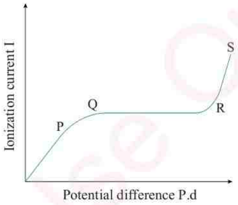


Figure 1.25 (b) Ionization curve

Explanation of the curve: On applying potential difference electrons and positive ions move slowly to the plate A (anode) and plate B (cathode), respectively; during this movement some ions recombine to form neutral atoms. Increase of the potential difference makes ions accelerate more quickly thus less opportunity for recombination to occur, resulting into an increase in ionization current as indicated by sensitive meter G.

As potential difference is further increased, it reaches a point Q where all the ions produced by the radiation reach the respective electrodes. In this region there is no recombination occurring and further increase of potential difference results into the current remaining more or less constant between Q and R. Thus ionization current reaches its saturation value and does not depend on the applied potential difference.

Beyond point R the original radiation ions are sufficiently accelerated by the large potential difference, forming new electrons and new positive ions by collision with the neutral gas molecules between the plates. Hence along RS each original ion-electron pair creates several other ion-electron pairs forming a highly ionized gas (plasma).

1.3.2 Optical spectra of gases

Optical spectra fall into two basic groups namely, absorption spectra and emission spectra. Absorption spectra are observed when part of the radiation emitted by a source of continuous spectra is absorbed by a material between the source and the observer. Absorption spectra can be line, band or continuous.

The emission spectra comprises of band spectra, line spectra and continuous spectra. In band spectra there are separate groups of lines known as bands, produced by gases and vapours whose molecules contain more than one atom (polyatomic vapours and gases such as O_2 and CO). Continuous spectra are produced by hot solids, liquids and high density gases. Line spectra are produced by monoatomic gas.

There are bright lines or bands on dark background produced when atoms of dilute gas in low pressure environment are excited and these excited electrons emit photons of specific wavelengths when they undergo transition to lower energy levels. Each element has a distinct set of electron energy levels and consequently result into photons of specific energies to be emitted. The spectrum produced is therefore unique for a given element. Observation of these emission lines allows the identification of the constituent elements in the gas. Emission spectra of different elements are shown in Figure 1.26.

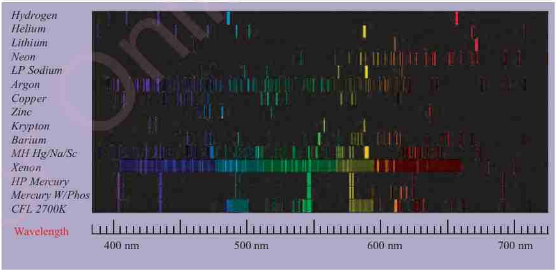


Figure 1.26 Emission spectra of different elements

To view and measure the atomic spectrum, the light emitted by the discharged gas is passed through a diffraction grating (or prism) in a spectrometer Figure 1.27 to disperse the light into its component wavelengths. Individual spectral lines (colours) can be viewed through the telescope of a spectrometer and each lines can be precisely located using the cross wires of the telescope. The angle at which the lines appear are read off a scale on the base of the spectrometer. For a diffraction grating, the angular location θ of a maxima for a particular spectral line of wavelength λ is given by:

$$m\lambda = d\sin\theta \qquad (1.39)$$

where d is the distance between the slits and m is the diffraction order.

From the measured angular location θ and the knowledge of d from the grating provided, the wavelength can be determined. If the wavelength λ for each line present in the spectra are determined, they can be compared with known spectrum for different gases.

Activity 1.3

Using a spectrometer and a diffraction grating, perform an experiment to determine the emission spectra of a gas discharge lamp (cadmium, mercury or sodium vapour source).

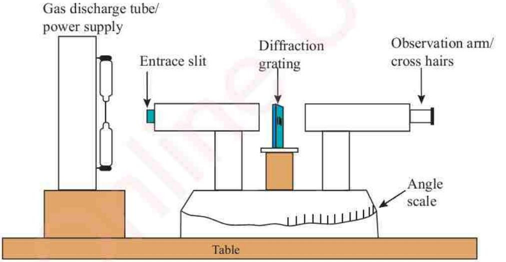


Figure 1.27 A spectrometer set up for measurement of emission spectral lines

1.3.3 Applications of conduction of electricity in gases

Current conduction in gases have wide and growing applications in daily life such as in fluorescent lighting, vapour lamps, surface treatment, flat plasma display panels for large area television screens and in research.

(a) Fluorescent lighting: A mercury vapour discharged tube gives out short-wave ultraviolet light which is itself invisible. When Ultraviolet light falls on the phosphor coating on the inside of the glass tube, the tube glows brilliantly in the visible wavelengths.

(b) Vapour lamps (Discharged lamps)

Passing an electric current through a gas at low pressure contained in a tube fitted with metal electrodes at each end, the gas will glow with a characteristic colour when a high voltage is applied across. The light output of a simple discharge lamp is in general very low and such lamps are commonly used for advertising signs. Small neon tubes are frequently used as indicator lamps. The colour of the light emitted depends upon the type of the gas used. The colour obtained from some of the gases and vapours commonly employed are listed in Table 1.2.

Table 1.2 Colours of light from discharge lamps for different gases and vapours

Gases/Vapour	Colours of light
Neon	Red
Hydrogen	Pink
Helium	Ivory
Nitrogen	Buff
Mercury (at low pressure)	Blue together with strong UV emissions
Mercury (at high pressure)	Bluish white with less UV emissions
Sodium	Yellow

The emission line spectra of some spectral lamps are shown in Table 1.3

Table 1. 3 Emission line spectra of some light sources

Light Source	Brightest lines in the spectrum of various gases						
Hydrogen gas	Color	Violet	Blue	Green	Yellow	Orange	Red
	Wavelength (nm) :	434	487				656
	:	410					
Helium gas	Color	Violet	Blue	Green	Yellow	Orange	Red
	Wavelength (nm) :	447	471	492	587	668	656
	:			502			
Neon gas	Color	Violet	Blue	Green	Yellow	Orange	Red
	Wavelength (nm) :				585	607	622
	:			540	588	616	627
Mercury gas	Color	Violet	Blue	Green	Yellow	Orange	Red
	Wavelength (nm) :	404	436	546	577		
	:				579		
Krypton gas	Color	Violet	Blue	Green	Yellow	Orange	Red
	Wavelength (nm) :	427		557	587		
	:	436					

Physics Form VI X.indd 27 10/11/2019 17:16

Exercise 1.2

1. Briefly explain conduction of electricity through gases. Discuss both ionized gases at ordinary pressure and discharged tube condition.
2. Discuss the successive steps in the appearance of the discharged tube as the gas pressure within the tube is steadily reduced.
3. Describe the possible optical spectra as related to the gas discharge.
4. Sketch and discuss the current voltage curve for electrical conduction in gases.
5. Explain at least two applications of conduction of electricity in gases.

1.4 Alternating current theory

When a resistor is connected across the terminals of a battery, current flows in the circuit. The current has a specific direction, it flows from the positive terminal to the negative terminal through an external load e.g. a resistor with a magnitude that remains constant. This is a direct current (*d.c.*). If the current in the resistor or in another element changes alternately and periodically, it is called an *alternating current (a.c.)*.

Alternating current is commonly applied at homes, offices, and factories that receive such energy from power companies. The flow of electrons alternates within the range of 50 to 60 revolutions per second (i.e. 50 Hz - 60 Hz). Alternating current is generated by an *a.c.* electric generator, which determines the frequency. In an *a.c.* the voltage can be readily changed to high or low values, therefore is more suitable for long-distance transmission than *d.c.* electricity because loss of power is less compared to the latter.

1.4.1 Alternating voltage and current

When a coil rotates in a magnetic field with uniform angular velocity ω, about an axis perpendicular to the magnetic field, an e.m.f. is induced in it. The magnitude of the induced *e.m.f.* (E) at any instant t, is given by:

$$E = E_o \sin\omega t \qquad (1.40)$$

where E_o is the maximum *e.m.f.* in the coil and ω is the angular frequency.

Equation (1.40) represents an *e.m.f.* which varies with time from zero to a maximum in one direction and again, from zero to a maximum in the opposite direction. During first half of the cycle, the *e.m.f.* is positive and during the second half of the cycle, the *e.m.f.* is negative as shown in Figure 1.28.

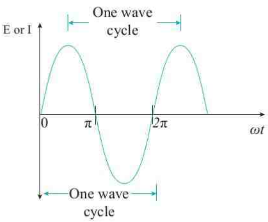


Figure 1.28 Variation of an alternating current or e.m.f. with time

Based on equation (1.40) the corresponding instantaneous value of current is given by:

$$I = \frac{E}{R} = \frac{E_o \sin \omega t}{R} \qquad (1.41)$$

$$I = I_o \sin \omega t$$

$$I = I_o \sin 2\pi f t \qquad (1.42)$$

where $I_o = \frac{E_o}{R}$ is known as the peak value current and f is the frequency, i.e. number of oscillations per second.

The angle $\omega t = 2\pi f t$ is known as instantaneous phase. This current is sinusoidal and is known as an alternating current. The variations of an alternating current or e.m.f. with time is shown in Figure 1.28.

The period of oscillation (T) is defined as the time for one complete cycle of oscillation and the frequency of oscillation f, i.e. the number of cycles per second, is given by:

$$f = \frac{1}{T} = \frac{\omega}{2\pi} \qquad (1.43)$$

1.4.2 Mean and root mean square (rms) value of alternating current and voltage

As already seen, the alternating current and *e.m.f.* vary in magnitude with time and change the direction periodically. The mean value of alternating current and *e.m.f.* for complete cycle are each zero. An ordinary ammeter or voltmeter will indicate this mean value. The mean value for the first half cycle is given by

$$\frac{\int_0^{\pi} E_o \sin \omega t \, d(\omega t)}{\int_0^{\pi} d(\omega t)} = \frac{2E_o}{\pi} \qquad (1.44)$$

For the second half cycle the mean value is given as in equation (1.44).

$$\frac{\int_{\pi}^{2\pi} E_o \sin \omega t \, d(\omega t)}{\int_{\pi}^{2\pi} d(\omega t)} = -\frac{2E_o}{\pi} \qquad (1.45)$$

Summation of cycles given by equations (1.44) and (1.45) gives;

$$\frac{2E_o}{\pi} - \frac{2E_o}{\pi} = 0 \qquad (1.46)$$

Therefore, for a complete cycle, the mean value is zero. Hence, an alternating current has to be independent of the direction of current. If the current (I) is flowing through a resistor R, then the heat developed is I^2R. Therefore, this principle is used to measure the current and voltage, whether I is positive or negative. The instruments based on this principle are known as hot wire instruments.

(a) Mean value of alternating current

The mean value of an alternating current is taken around half of the cycle, it is because the area of the positive half cycle is equal to that of the negative half cycle. However, one can find the mean value of alternating current over any half cycle.

The half cycle mean value of an *a.c.* is equivalent to that value of a constant direct current which would send the same amount of charge *(Q)* through a circuit for half of the time period of an *a.c.* as is sent by an *a.c.* through the same circuit and time. The half cycle mean value of an alternating current is denoted by symbol I_m.

The instantaneous value of an *a.c.* is given by the relation:

$I = I_o \sin\omega t$

Assuming $I = I_o \sin\omega t$ remains constant for a very small time (dt) then the small amount of charge sent by an *a.c.* during this time interval (dt) is given by the relation:

$$dQ = Idt = I_o \sin\omega t\, dt$$

If Q is the total charge in the positive half cycle of an alternating current, then:

$$Q = \int_0^{\frac{T}{2}} I\,dt = \int_0^{\frac{T}{2}} I_o \sin\omega t\,dt$$

$$Q = I_o \left|\frac{-\cos\omega t}{\omega}\right|_0^{\frac{T}{2}} = -\frac{I_o}{\omega}\left[\cos\frac{\omega T}{2} - \cos 0^0\right]$$

$$Q = -\frac{I_o}{\omega}\left[\cos\pi - \cos 0^0\right]$$

$$Q = \frac{2I_o}{\omega} \qquad (1.47)$$

If I_m is the half cycle average or mean value of positive half cycle of an *a.c.* then by definition:

$$Q = I_m \times \frac{T}{2} \qquad (1.48)$$

By equating equations (1.47) and (1.48) we get:

$$I_m \times \frac{T}{2} = \frac{2I_o}{\omega}$$

$$I_m \times \frac{T}{2} = 2I_o \times \frac{T}{2\pi}$$

$$I_m = \frac{2}{\pi} I_o = 0.637 I_o$$

$$I_m = 0.637 I_o \qquad (1.49)$$

Therefore, the half cycle mean value is + 0.637 I_o for positive half cycle and for negative half cycle is – 0.637 I_o, where I_0 is the peak value of an alternating current.

(b) Mean value of alternating voltage

The half cycle mean of an alternating *e.m.f.* is that value of an alternating e.m.f. which would send the same amount of charge (Q) through a circuit for half of the time of alternating *e.m.f.* as sent by the steady direct e.m.f. through the same circuit at the same time. The half cycle mean value of an alternating e.m.f. is represented by symbol(E_m).

The instantaneous value of an alternating e.m.f. is given by the relation;

$E = E_o \sin\omega t$

Assuming this alternating *e.m.f.* is applied to a circuit of resistance R, then by Ohm's law the instantaneous value of alternating current is:

$$I = \frac{E}{R} = \frac{E_o \sin\omega t}{R} = \frac{E_o}{R}\sin\omega t$$

If the said current remains constant for a very small time interval (dt), then small amount of charge sent by alternating e.m.f. during this time (dt) is given by:

$$dQ = Idt = \frac{E_o}{R}\sin\omega t\,dt$$

If Q is the total charge in the positive half cycle of an alternating *e.m.f.*, then;

$$Q = \int_0^{\frac{T}{2}} \frac{E_o}{R}\sin\omega t\,dt = \frac{E_o}{R}\left[-\frac{\cos\omega t}{\omega}\right]_0^{\frac{T}{2}}$$

$$Q = -\frac{Eo}{\omega R}\left[\cos\frac{\omega T}{2} - \cos 0\right]$$

$$Q = \frac{2E_o}{\omega R} \qquad (1.50)$$

If E_m is the half cycle mean value of positive half cycle of an alternating e.m.f. then by definition:

$$Q = \frac{E_m}{R} \times \frac{T}{2} \qquad (1.51)$$

By equating equations (1.50) and (1.51) we obtain:

$$\frac{E_m}{R} \times \frac{T}{2} = \frac{2E_o}{\omega R}$$

$$E_m \times \frac{T}{2} = 2E_o \times \frac{T}{2\pi}$$

$$E_m = \frac{2}{\pi} E_o = 0.637 E_o$$

$$E_m = 0.637 E_o \qquad (1.52)$$

Therefore, the half cycle mean value is + 0.637 E_o for positive half cycle and for negative half cycle is – 0.637 E_o where E_0 is the peak value of an alternating (*e.m.f.*).

(c) Root mean square (rms) value of alternating current

To specify an alternating current the mean value cannot be used because its value is zero over one complete cycle. Also, the mean value cannot be used for power calculation. To measure the effectiveness of an alternating current one must measure it in terms of equivalent direct current that would do work or produce heat at the same average rate as an alternating current under the same conditions. This equivalent direct current is known as the root mean square (rms) value of an alternating current.

The root mean square (rms) value (also called effective value) of an alternating current is the alternating current which when flowing through a given resistor for a given time converts electrical energy to other forms of energy, e.g. heat, as produced by a steady direct current when flowing through the same resistor for the same time. The root mean square (*rms*) of an alternating current is denoted by symbol I_{rms}. If the alternating current I, flows through a resistance (*R*) for a small time (*dt*), then the small amount of heat generated (*dH*) is given by:

$$dH = I^2 R dt$$

But, $I = I_o \sin \omega t$

Thus,

$$dH = \left(I_o \sin \omega t\right)^2 R dt = I_0^2 R \sin^2 \omega t\, dt$$

For one complete cycle of an alternating current, the heat generated in the resistor of resistance *R* can be calculated as:

$$\int_0^H dH = \int_0^T I_0^2 R \sin^2 \omega t\, dt$$

$$H = I_0^2 R \int_0^T \sin^2 \omega t\, dt$$

$$H = I_0^2 R \int_0^T \left(\frac{1 - \cos 2\omega t}{2}\right) dt$$

$$H = \frac{I_0^2 R}{2} \left| t - \frac{\sin 2\omega t}{2\omega} \right|_0^T$$

$$H = \frac{I_0^2 R}{2} \left[T - \frac{\sin 2 \times 2\pi}{2\omega} \right]$$

$$H = \frac{I_0^2 RT}{2} \qquad (1.53)$$

Generally, the heat energy generated in

a resistor of resistance R for a complete cycle is given as:

$$H = I_{rms}^2 RT \tag{1.54}$$

Hence, equating equations (1.53) and (1.54) one can obtain:

$$I_{rms}^2 RT = \frac{I_o^2 RT}{2}$$

$$I_{rms} = \frac{I_o}{\sqrt{2}}$$

$$I_{rms} = 0.707 I_o \tag{1.55}$$

Therefore, the root mean square value of an alternating current is $0.707 I_o$ where I_0 is the peak value of an altenating current. It should be noted that, root mean square value of an alternating current is the same for both half and full cycle.

(d) Root mean square (rms) value of alternating e.m.f.

The root mean square value of alternating electromotive force is that alternating voltage which on application to a resistor of resistance R for a given time, will produce the same amount of heat energy as if a constant direct *e.m.f.* is applied to the same resistor under the same conditions. This is sometimes referred as virtual value of alternating *e.m.f.* and is denoted as E_{rms}.

The instantaneous value of an alternating e.m.f. is given by the relation:

$$E = E_o \sin \omega t$$

If the alternating voltage is applied to a resistor of resistance R for a small time interval dt, then the small amount of heat created is given by:

$$dH = \frac{E^2}{R} dt = \frac{(E_o \sin \omega t)^2}{R} dt$$

$$dH = \frac{E_o^2}{R} \sin^2 \omega t \, dt$$

For one complete cycle, the amount of heat produced in a resistance (R) is:

$$\int_0^H dH = \int_0^T \frac{E_o^2}{R} \sin^2 \omega t \, dt$$

$$H = \frac{E_o^2}{R} \int_0^T \sin^2 \omega t \, dt$$

$$H = \frac{E_o^2}{R} \int_0^T \frac{(1 - \cos 2\omega t)}{2} dt$$

$$H = \frac{E_0^2 T}{2R} \tag{1.56}$$

Assuming that E_{rms} is applied to the resistor of resistance R for the same time T, then the heat energy generated is given as:

$$H = \frac{E_{rms}^2 T}{R} \tag{1.57}$$

Thus, by equating equations (1.56) and (1.57) one can have;

$$\frac{E_{rms}^2 T}{R} = \frac{E_0^2 T}{2R}$$

$$E^2{}_{rms} = \frac{E_o{}^2}{2}$$

$$E_{rms} = \frac{E_o}{\sqrt{2}}$$

$$E_{rms} = 0.707 E_o \tag{1.58}$$

Thus, the root mean square value of an alternating e.m.f. is $0.707 E_o$ where E_0 is the peak value of an alternating *e.m.f.*. It is essentially to note that, root mean square value of an alternating *e.m.f.* is the same for both half and full cycle.

Activity 1.4

You are provided with *a.c.* and *d.c.* ammeters, variable resistor, connecting wires, 2.5 V lamp, two way switch, a 3 V battery, and a.c. power supply of about 2 V. Connect the circuit as shown in Figure 1.29

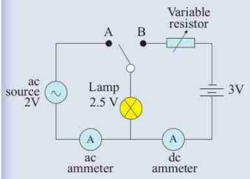


Figure 1.29 A two way switch circuit.

(a) Record the value of *d.c.* current when the switch is closed to point B and adjust the rheostat until there is no visible change in the brightness of the lamp when a switch is moved from one point to the other.

(b) Repeat the experiment using at least five different values of alternating voltage. The two values recorded should be the same in each case. Why?

(c) Repeat the experiment by replacing the lamp with CRO. Record the *d.c.* voltage V_{rms} when the switch is in point B and the peak to peak voltage, $2V_o$ i.e. vertical height from top to bottom on trace, when the switch is in point A.

(d) Tabulate the results and calculate the peak voltage V_o in the a.c. circuit for each case.

(e) Draw a graph of V_o against V_{rms}. Find the gradient of the graph. What does this gradient represent?

Example 1.10

An *a.c.* circuit of frequency 50 Hz has a maximum current of 100 A.

(a) Write down the equation for the instantaneous values of the current.

(b) Find the instantaneous values after 1/180 seconds.

Solution

(a) Instantaneous values of current

$$I = I_o \sin \omega t = I_o \sin 2\pi f$$

$$t = 100 \sin\left(2\pi \times 50\,Hz\right)t$$

$$I = 100 \sin 100\pi t$$

(b) Given seconds then;

$$I = 100\,\text{A} \times \sin\left(100\pi \times \frac{1}{180}\text{s}\right)$$

$$= 100 \sin \frac{5}{9}\pi = 98.48\ \text{A}$$

Example 1.11

A 60 ohms electric lamp is connected to a 240 V, 50 Hz supply. Determine:

(a) the effective voltage; and

(b) the peak value voltage.

Solution

(a) As the power supply is 240 V *a.c.* therefore, the effective voltage is 240 V

(b) The peak value of voltage is given as $E_{rms} = \dfrac{E_o}{\sqrt{2}}$

Therefore

$$E_o = \text{E}_{\text{rms}} \times \sqrt{2} = 240\,\text{V} \times 1.41 = 339\,\text{V}$$

Example 1.12

In an electric circuits, an applied alternating voltage is $E = 30\sin(300t)\ V$ and the current in the circuit is $I = \sin(300 + \pi/3)\ A$. Calculate:

(a) the frequency of the applied voltage;
(b) rms voltage and current.

Solution

(a) Comparing the given equation and the standard equation we have

$E = E_o \sin\omega t$ and $E = 30\sin(300t)$

$\omega t = 300t$ and $2\pi f = 300$

Therefore, $f = \dfrac{300}{2\pi} = 47.75\text{Hz}$

(b) By comparison, the peak voltage value is 30 volts, thus

$E_{rms} = \dfrac{E_o}{\sqrt{2}} = \dfrac{30\text{ V}}{\sqrt{2}} = 21.21\text{ V}$ and

$I_{rms} = \dfrac{I_o}{\sqrt{2}} = \dfrac{1\text{A}}{\sqrt{2}} = 0.707\text{ A}$

Example 1.13

The peak value of an alternating current is 5 A and its frequency is 60 Hz.

(a) Find its rms value.
(b) How long will the current take to reach the peak value starting from zero?

Solution

(a) $I_{rms} = \dfrac{I}{\sqrt{2}} = \dfrac{5\,A}{\sqrt{2}} = 3.54\text{ A}$

(b) Time period T is

$$T = \frac{1}{f} = \frac{1}{60}\text{s} = 0.0167\text{ s}$$

The current takes one fourth period to reach the peak value starting from zero. Thus, the time required is

$$t = \frac{T}{4} = \frac{1}{4\times 60}$$

$$= \frac{1}{240}\text{s} = 4.1667\times 10^{-3}\text{s}$$

1.4.3 Alternating Current Circuits

(a) Circuit with resistor only

Figure 1.30 (a) shows an alternating current source connected to a resistor.

Such a circuit is sometimes called a purely resistive circuit as it has only that electrical element.

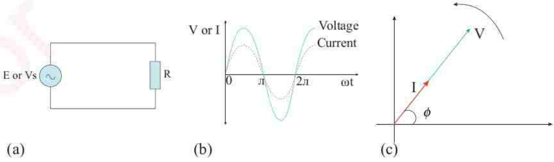


Figure 1.30 A purely resistive circuit and its phasor diagrams

If the current at time t is I, Kirchhoff's loop law gives:

$$E_o \sin \omega t = IR$$

or

$$I = \frac{E_o \sin \omega t}{R} = \frac{E_o}{R} \sin \omega t = I_o \sin \omega t$$

$$I = I_o \sin \omega t \qquad (1.59)$$

where $I_o = \dfrac{E_o}{R}$

The voltage and current are always in phase i.e. the corresponding variation in I and E occur simultaneously, that is, both I and E are zero or maximum or minimum at the same instant Figure 1.30 (b) and (c). Instantaneous power consumed is defined as the product of current and voltage at that instant.

$$P = IE$$

$$P = I_o \sin \omega t \times E_o \sin \omega t$$

$$P = I_o E_o \sin^2(\omega t)$$

But for a complete cycle the average value of $\sin^2(\omega t)$ is $\frac{1}{2}$. The average power P_{av} is therefore:

$$P_{av} = \frac{1}{2} \times I_o E_o = \frac{I_o}{\sqrt{2}} \frac{E_o}{\sqrt{2}}$$

$$P_{av} = \frac{I_o}{\sqrt{2}} \frac{E_o}{\sqrt{2}} = I_{rms} E_{rms} \qquad (1.60)$$

Power consumed by the resistor in an *a.c.* circuit is equal to the product of I_{rms} and E_{rms}.

Example 1.14

What will be the resistance of a resistor of 100 Ω at frequencies of 60 Hz and 100 kHz?

Solution

The resistance of the resistor for *a.c.* is the same, and equal to 100 Ω for the two frequencies 60 Hz and 100 kHz.

(b) Circuit with inductor only

If an alternating voltage is applied across a pure inductor, a back *e.m.f.* $(-L\frac{dI}{dt})$ is induced in the inductor due to its self-inductance. $\frac{dI}{dt}$ is the rate of change of current and I is the current in the circuit at any instant. The negative sign implies that induced e.m.f. opposes the change in current. Figure 1.31(a) shows an inductor connected to an external source; such a circuit is called a purely inductive circuit.

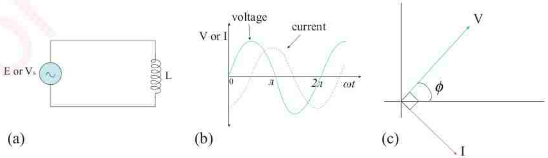


Figure 1.31 A purely inductive circuit and its corresponding phasor diagrams

The applied voltage is equal and opposite to the induced e.m.f. at any instant. Thus:

$$E_o \sin\omega t - L\frac{dI}{dt} = 0$$

$$dI = \frac{E_o}{L}\sin(\omega t)dt$$

Integrating both sides

$$\int dI = \int \frac{E_o}{L}\sin(\omega t)dt$$

$$I = -\frac{E_o}{\omega L}\cos\omega t + c$$

where c is a constant of integration. The average of $\cos\omega t$ over one time period T is zero and since the expression has a cos term, c must be zero, so that I, can be written as:

$$I = -\frac{E_o}{\omega L}\cos\omega t \qquad (1.61)$$

From trigonometric relationships:

$-\cos\omega t = \sin(\omega t - \frac{\pi}{2})$. Hence,

$$I = I_o \sin\left(\omega t - \frac{\pi}{2}\right) \qquad (1.62)$$

where $I_o = \dfrac{E_o}{\omega L}$

Equation (1.62) shows that, the phase of the current is $\pi/2$ behind that of the *e.m.f.* as indicated in Figures 1.31 (b) and 1.31(c). This observation can also be stated as follows: The voltage *leads* the current by $\pi/2$ or the current lags behind the voltage by $\pi/2$.

From equations (1.61) and (1.62). Let

$$X_L = \omega L = 2\pi f L \qquad (1.63)$$

X_L is called the inductive reactance, it is found to be directly proportional to the frequency of the source at a constant inductance as illustrated in Figure 1.32 (a). Also, X_L is directly proportional to the inductance at a constant frequency as shown in Figure 1.32 (b).

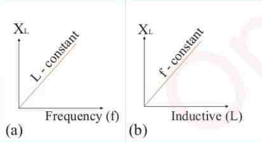


Figure 1.32 Variation of inductive reactance with frequency and inductance

From the definition of instantaneous power consumed in the inductor

$$P = IE$$

$$P = -I_o \cos\omega t \times E_o \sin\omega t$$

$$P = I_o E_o \sin\omega t \cos\omega t$$

$$P = -\frac{1}{2} \times I_o E_o \sin 2\omega t$$

But for a complete cycle an average value of $\sin 2\omega t$ is zero, $P_{av} = 0$.
Hence the average power consumed by pure inductor is zero.

Example 1.15

An inductor of inductance L= 200 mH is connected to an *a.c.* source of *e.m.f.* 210 V and frequency of 50 Hz. Calculate:

(a) reactance of the inductor;
(b) peak current; and
(c) instantaneous voltage of the source when the current is at its peak value.

Solution

(a) The reactance of the inductor is

$$X_L = \omega L = 2\times\pi\times 50\,\text{Hz}\times 200\times 10^{-3}\,\text{H}$$
$$= 62.8\,\Omega$$

Therefore, the reactance of the inductor is $X_L = 62.8\,\Omega$

(b) The peak current (I_o) is

$$I_o = \frac{E_o}{X_L} = \frac{210\,\text{V}\sqrt{2}}{62.8\,\Omega} = 4.73\,\text{A}$$

The peak current (I_o) is 4.73A

(c) As the current lags behind the voltage by π/2, the voltage is zero when the current has its peak value.

Example 1.16

The current flowing through a 0.6 H inductor changes sinusoidally with an amplitude of 2.0 A and a frequency of 50 Hz. Calculate the root mean square potential difference across the terminals of the inductor.

Solution

From $I = I_0 \sin\omega t$

Induced e.m.f. across the inductor

$$E = L\frac{d}{dt}\left(I_o \sin\omega t\right) = I_o\omega L\cos\omega t$$

Therefore,

E_o = amplitude of the potential difference

$$= I_o\omega L = 2\pi f L I_o$$

$$E_o = 2\pi\times 50\,\text{Hz}\times 0.6\,H\times 2\,\text{A} = 376.99\text{V}$$

$$E_{rms} = \frac{E_o}{\sqrt{2}} = \frac{376.99\text{V}}{\sqrt{2}} = 266.57\text{V}$$

Thus, $E_{rms} = 266.57\text{V}$

(c) Circuit with capacitor only

Alternating current can flow through a resistor but it is not obvious at first that it can flow through the capacitor. If an alternating current is applied to a capacitor, the capacitor is charged first in one direction and then in the opposite direction. The results is that electrons move back and forth round the circuit connecting the plates, hence, establishing alternating current.

Figure 1.33 (a) shows an electric circuit with *a.c.* power source connected across a capacitor, such a circuit is known as a purely capacitive circuit.

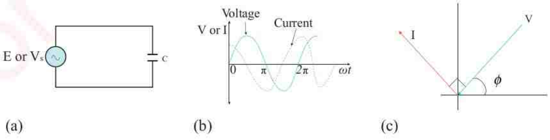


Figure 1.33 A purely capacitive circuit and its corresponding phasor diagrams

Suppose the charge on the capacitor is Q at time t, then the charge accumulating on the capacitor plates contributes to the current I as:

$$Idt = dQ \quad \text{or} \quad I = \frac{dQ}{dt}$$

Using Kirchhoff's voltage law we get;

$$E_o \sin \omega t = \frac{Q}{C} \quad \text{or} \quad Q = C\, E_o \sin \omega t$$

But;

$$I = \frac{dQ}{dt} = \frac{d}{dt}\left(CE_o \sin \omega t\right) = CE_o \omega \cos \omega t$$

$$I = I_o \cos \omega t$$

Using trigonometric relations

$$I = I_o \sin\left(\omega t + \frac{\pi}{2}\right) \qquad (1.64)$$

where, $I = I_o \sin\left(\omega t + \frac{\pi}{2}\right)$

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C}$$

The phase difference between the *e.m.f.* and the current is $\pi/2$. Based on the equations (1.40) and (1.64), or Figures 1.33 (b) and 1.33 (c) the current leads the e.m.f. by $\pi/2$, implying that when the *e.m.f.* is zero the current has a maximum magnitude, and vice versa. $X_C = \frac{1}{\omega c}$ is called the capacitive reactance of the capacitor and its unit is ohm (Ω). This equation is a function of the capacitance of the capacitor and the frequency of the *a.c.* source.

The capacitive reactance is found to be inversely proportional to the frequency at a constant capacitance, and inversely proportional to the capacitance at a constant frequency as shown in Figures 1.34 (a) and 1.34 (b), respectively.

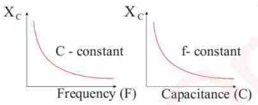


Figure 1.34 Variation of capacitive reactance of the capacitor

From the definition of instantaneous power consumed in the capacitor,

$$P = IE$$

$$P = I_o \cos \omega t \times E_o \sin \omega t$$

$$P = I_o E_o \sin \omega t \cos \omega t$$

$$P = \frac{1}{2} \times I_o E_o \sin 2\omega t$$

But for a complete cycle an average value of $\sin 2\omega t$ is zero, $P_{av} = 0$. Hence the average power consumed by pure capacitor is zero.

No power is stored by either capacitor or inductor over a complete cycle, and thus the average power stored by either capacitors or inductors is zero. A capacitor stores energy during the first half cycle, that the electric charge on either plates is increasing, but losses the energy to the source in the other half cycle as the charge falls to zero. Similarly, an inductor store energy as its magnetic field rises but losses the energy to the supply when the field fall to zero.

Example 1.17

What will be the reactance of a capacitor of 100 μF at frequencies of 60 Hz and 100 kHz?

Solution

At 60 Hz

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi fC}$$

$$= \frac{1}{2\times\pi\times 60\ \text{Hz}\times 10^{-4}\ \text{F}} = 26.53\ \Omega$$

The reactance at 60 Hz is $26.53\ \Omega$

At 100 kHz

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi fC}$$

$$= \frac{1}{2\times\pi\times 100\times 10^{3}\ \text{Hz}\times 10^{-4}\text{F}}$$

$$= 1.59\times 10^{-2}\ \Omega$$

The reactance at 100 kHz is $1.59\times 10^{-2}\ \Omega$

Example 1.18

A radio circuit has a capacitor of $2\ \propto F$ with the frequency 1 kH and the root mean square current flowing is 4 mA.

(a) Calculate the potential difference (p.d.) across the capacitor.

(b) What is the current flowing when root mean square voltage of 10 V, 50 Hz is connected to this capacitor?

Solution

(a) From $X_C = \frac{1}{\omega C} = \frac{1}{2\pi fC}$

$$= \frac{1}{2\times\pi\times 1000\ \text{Hz}\times 2\times 10^{-6}\text{F}}$$

$$= 79.58\ \Omega$$

But

$$V_C = IX_C = 4\times 10^{-3} A\times 79.58\ \Omega = 0.32\ \text{V}$$

The potential difference across the capacitor is 0.32 V

(b) $X_C \propto \frac{1}{f}$, since the capacitance is constant

$$\frac{X_{C1}}{X_{C2}} = \frac{f_2}{f_1} \Rightarrow X_{C2} = \frac{f_1\times X_{C1}}{f_2}$$

$$= \frac{1000\ \text{Hz}\times 79.58\Omega}{50\ \text{Hz}} = 1591.6\ \Omega$$

From $I_{rms} = \frac{V_{rms}}{X_{C2}} = \frac{10\ \text{V}}{1591.6\ \Omega}$

$$= 6.28\times 10^{-3}\ \text{A} = 6.28\ \text{mA}$$

The current flowing is 6.28mA

Example 1.19

A capacitor of 10 μF is connected to the terminals of 60 Hz *a.c.* source whose root mean square voltage is 200 V. Calculate the capacitive reactance and the root mean square current in the circuit.

Solution

The capacitive reactance is

$$X_C = \frac{1}{\omega C}$$

$$= \frac{1}{2\times\pi\times 60\ \text{Hz}\times 10\times 10^{-6}\text{F}} = 265.26\Omega$$

The root mean square current is

$265.26\ \Omega$

$$I_{rms} = \frac{V_s}{X_c} = \frac{200.00\,\text{V}}{265.26\,\Omega} = 0.75\text{ A}$$

The root mean square current is 0.75 A

1.4.4 More a.c. circuits

When an *a.c.* source is connected in a circuit with a resistor, capacitor or inductor, the current varies initially in a complex way but after sufficient time, a sinusoidal current persists in the circuit. This steady state current has a frequency equal to that of the source and may have a phase which is different from that of the source. This current may be obtained by the vector method described in the following section.

(a) Resistor-Capacitor (RC) circuit in series

An RC circuit consists of a resistor and capacitor connected in series across an *a.c.* source as depicted in Figure 1.35(a).

The total opposition offered to current flow by the combination effect of resistor R and capacitor C is called impedance, denoted by the letter Z and its SI unit is Ohm (Ω).

The impedance (Z) of the circuit can be deduced using a vector diagram. Let V_S, V_R and V_C be source voltage, potential difference across resistor and potential difference across capacitor respectively. It is known that for the resistor, the current and voltage are in phase while for a capacitor, the current leads the voltage by $\pi/2$. Therefore, V_R leads V_C by $\pi/2$ as indicated in Figure 1.35(b).

Using Pythagoras theorem,

$$V_s^2 = V_R^2 + V_c^2 \quad \Rightarrow \quad V_s = \sqrt{(IR)^2 + (IX_c)^2}$$

but, $X_C = \dfrac{1}{\omega c}$

$$\frac{V_s}{I} = \sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}$$

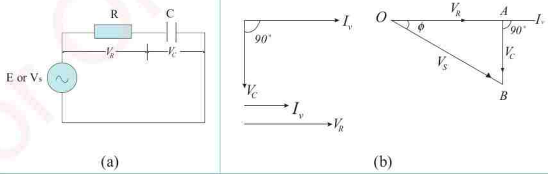


Figure 1.35 An alternating RC circuit and a vector diagram for RC circuit

$$\therefore I = \frac{V_s}{\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}}$$

Let $\frac{V_S}{I} = Z$, which is the impedance of the circuit, so that

$$Z = \sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2} \qquad (1.65)$$

From Figure 1.35 (b) the phase angle ϕ is obtained as follows:

$$\tan\phi = \frac{V_C}{V_R} = \left(\frac{IX_C}{IR}\right) = \frac{X_C}{R} = \frac{1}{\omega CR}$$

$$\phi = \tan^{-1}\left(\frac{1}{\omega CR}\right) = \tan^{-1}\left(\frac{X_C}{R}\right) \qquad (1.66)$$

Example 1.20

A 100 $\propto F$ capacitor is connected in series with a 5V, 0.30 A lamp, and a 50 Hz supply. Determine:

(a) V_{rms} of the supply to light the lamp to its normal brightness:

(b) potential difference across the capacitor and the resistor respectively.

Solution

(a) $X_C = \frac{1}{2\pi fC}$

$= \frac{1}{2\pi \times 100 \times 10^{-6}\ \text{F} \times 50\ \text{Hz}}$

$= 31.83\Omega$

The resistance R, of lamp

$$R = \frac{V}{I} = \frac{5.0\ \text{V}}{0.3\text{A}} = 16.67\Omega$$

The impedance Z of a circuit is

$$Z = \sqrt{R^2 + X_C^2}$$

$$= \sqrt{(16.67\Omega)^2 + (31.83\Omega)^2} = 35.93\Omega$$

The applied V_{rms} to cause I_{rms} of 0.3A to flow through the circuit,

$$V_{rms} = IZ = 0.3\text{A} \times 35.93\Omega = 10.78\ \text{V}$$

(b) The potential difference across the capacitor,

$$V_C = IX_C = 0.3\,A \times 31.83\ \Omega = 9.55\ \text{V}$$

The *p.d.* across the resistor,

$$V_R = IR = 0.3\text{A} \times 16.67\ \Omega = 5.00\ \text{V}$$

$V_c + V_R$ is always greater than the V applied. Justify.

Example 1.21

Calculate the capacitance of the capacitor which must be placed in series with a 5 Ω resistor and inductance of 10 mH to bring the current in phase with the voltage, for the frequency of 50 Hz.

Solution

The current and the voltage are in phase when;

$$X_C = X_L \Rightarrow \frac{1}{\omega C} = \omega L$$

$$C = \frac{1}{\omega^2 L}$$

$$= \frac{1}{(2\pi \times 50\ \text{Hz})^2 \times 10 \times 10^{-3}\ \text{H}}$$

$$= 1.013 \times 10^{-3}\ \text{F}$$

Therefore, the capacitance is 1.013 mF

(b) Inductor-Resistor (LR) circuit in series

The circuit in Figure 1.36(a) shows an inductor and a resistor connected in series to an alternating source V_s. As in the case of RC circuit, the current through an LR circuit depends on the values of both components and the frequency of the source.

Figure 1.36 (b) shows a vector diagram of an LR circuit. Assuming that the voltages across the resistor and the inductor are V_R and V_L respectively, and for the resistor, voltage and current are in phase while for the inductor the voltage leads the current by $\pi/2$.

By Pythagoras theorem

$$V_S = \sqrt{V_R^2 + V_L^2}$$

Since, $X_L = \omega L$, then

$$I = \frac{V_s}{\sqrt{R^2 + X_L^{\ 2}}} = \frac{V_s}{\sqrt{R^2 + (\omega L)^2}}$$

Again the impedance is given by:

$$Z = \frac{V_s}{I} = \sqrt{R^2 + X_L^2} = \sqrt{R^2 + (\omega L)^2}$$

$$Z = \sqrt{R^2 + (\omega L)^2} \qquad (1.67)$$

The phase angle ϕ can be expressed as follows:

$$\phi = \tan^{-1}\left(\frac{V_L}{V_s}\right) = \tan^{-1}\left(\frac{\omega L}{R}\right)$$

$$= \tan^{-1}\left(\frac{X_L}{R}\right) \qquad (1.68)$$

Example 1.22

A coil having a resistance of 10 Ω and inductance of 21 mH is connected to a 240 V, 60 Hz supply. Use this information to determine:

(a) current;
(b) phase angle;
(c) power factor ;
(d) power consumed ; and
(e) the time lag between the voltage maximum and the current maximum.

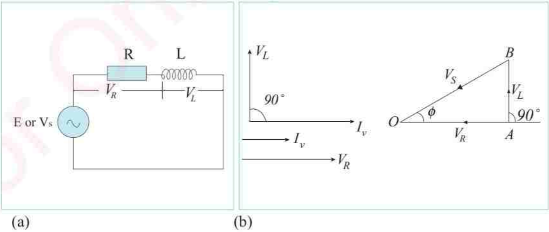


Figure 1.36 An alternating LR circuit and a vector diagram for LR circuit

Solution

(a) From $X_L = \omega L = 2\pi fL$

$$= 2\pi \times 60\,\text{Hz} \times 21 \times 10^{-3}\text{H} = 7.92\,\Omega$$

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(10\,\Omega)^2 + (7.92\,\Omega)^2}$$

$$= 12.76\,\Omega$$

$$I = \frac{V}{Z} = \frac{240\,V}{12.76\,\Omega} = 18.81\text{ A}$$

(b) Phase angle

$$\phi = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{7.92\,\Omega}{10.00\,\Omega}\right) = 38.4^\circ$$

(c) Power factor

$$= \cos\phi = \cos 38.4^\circ = 0.784$$

(d) Power consumed

$$P = IV\cos\phi = 18.81\text{A} \times 240\,\text{V} \times 0.784$$

$$= 3539\,\text{W} = 3.539\,\text{kW}$$

Thus, power consumed is 3.539 kW

(e) Time lag

$$\phi = \omega t \Rightarrow t = \frac{\phi}{\omega} = \frac{\phi}{2\pi f}$$

$$= \frac{2\pi \times 38.4}{360} \times \frac{1}{2\pi \times 60}$$

$$= 1.8 \times 10^{-3}\text{s} = 1.8\,\text{ms}$$

Therefore, the time lag is $1.8\,\text{ms}$

Example 1.23

When 120 V *d.c.* are applied across a coil, a current of 1.25 A flow through it. When 120 V *a.c.* of 50 Hz are applied to the same coil only 0.5 A flows. Determine:

(a) resistance;

(b) impedance; and

(c) inductance of the coil.

Solution

(a) When the *d.c.* is applied the resistance can be calculated as follows;

$$V = IR$$

$$R = \frac{V}{I} = \frac{120.0\,\text{V}}{1.25\,\text{A}} = 96\,\Omega$$

The resistance is $96\,\Omega$

(b) When the *a.c.* is applied the impedance can be calculated as follows:

$$Z = \frac{V}{I} = \frac{120\,\text{V}}{0.50\,\text{A}} = 240\,\Omega$$

The impedance is $240\,\Omega$

(c) $Z = \sqrt{R^2 + (\omega L)^2}$

$$\omega L = \sqrt{Z^2 - R^2}$$

$$L = \frac{\sqrt{Z^2 - R^2}}{\omega}$$

$$= \frac{\sqrt{(240\,\Omega)^2 - (96\,\Omega)^2}}{2\pi \times 50\text{ Hz}} = 0.70$$

The inductance of the coil is $0.70\,\text{H}$

(c) Resistor, Inductor and Capacitor (RLC) circuit in series

Figure 1.37 (a) shows RLC circuit in which a resistor, an inductor, and a capacitor are connected in series to an alternating current source. As the elements are in series, the current in the circuit is the same at any instant. Therefore, the current at all components has the same phase and amplitude.

The potential difference across each component has a different phase and amplitude. The *p.d.* across the resistor is in phase with the current, while the *p.d.* across the capacitor lags the current by $\pi/2$ and that across the inductor leads the current by $\pi/2$.

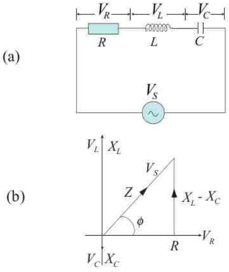


Figure 1.37 RLC series circuit and vector diagram

The instantaneous voltages across the three circuit elements is such that:

$$V_S = V_R + V_C + V_L$$

A vector diagram for the circuit is illustrated in Figure 1.37 (b). Using these vector relationship, the following cases can be observed.

Case I. If the circuit is net inductive V_L i.e. is greater than V_C then net potential difference across the circuit is obtained by Pythagoras's theorem,

$$V_S^2 = V_R^2 + \left(V_L - V_C\right)^2$$

$$V_S^2 = (IR)^2 + \left(IX_L - IX_C\right)^2$$

$$\frac{V_S^2}{I^2} = (R)^2 + \left(X_L - X_C\right)^2$$

The impedance Z can then be expressed as:

$$\frac{V^2}{I^2} = Z^2 = (R)^2 + \left(X_L - X_C\right)^2$$

$$Z = \sqrt{R^2 + \left(X_L - X_C\right)^2}$$

$$= \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \qquad (1.69)$$

When V_L is greater than V_C the phase angle is positive, implying that the current lags the applied voltage. When V_C is greater than V_L, the phase angle is negative, indicating that the current leads the applied voltage. The corresponding phase angle (ϕ) of the circuit can be expressed as:

$$\phi = \tan^{-1}\left(\frac{\left(X_L - X_C\right)}{R}\right)$$

$$= \tan^{-1}\left(\frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}\right) \qquad (1.70)$$

Case II. If the circuit is net capacitive i.e. V_C is greater than V_L then net potential difference across the circuit is given as:

$$V_S^2 = V_R^2 + \left(V_C - V_L\right)^2$$

$$\frac{V_S^2}{I^2} = \left(R\right)^2 + \left(X_C - X_L\right)^2$$

The impedance Z can be expressed as

$$\frac{V_S^2}{I^2} = Z^2 = (R)^2 + \left(X_c - X_L\right)^2$$

$$Z = \sqrt{R^2 + \left(X_C - X_L\right)^2}$$

$$= \sqrt{R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2} \quad (1.71)$$

The corresponding phase angle (ϕ) of the circuit can be expressed as:

$$\phi = \tan^{-1}\left(\frac{\left(X_C - X_L\right)}{R}\right)$$

$$= \tan^{-1}\left(\frac{\left(\frac{1}{\omega C} - \omega L\right)}{R}\right) \quad (1.72)$$

Case III. If X_C is equal to X_L in both cases (I) and (II) then;

$$Z = \sqrt{R^2 + \left(X_C - X_L\right)^2}$$

The total impedance reduces to $Z = R$. This shows that, the phase angle is zero, the impedance (Z) is minimum, current is maximum thus, the circuit is purely resistive.

$$X_L = X_C \Rightarrow X_L - X_C = 0$$

$$\omega L = \frac{1}{\omega C} \Rightarrow \omega^2 = \frac{1}{LC} \text{ where } \omega = 2\pi f$$

$$f_o = \frac{1}{2\pi\sqrt{LC}} \quad (1.73)$$

The resonant frequency $\left(f_o\right)$ is a frequency attained when the applied voltage produces maximum current in the RLC circuit. Figure 1.38 (a) shows that the resistance R is independent of frequency and is thus represented by a horizontal line parallel to the frequency axis.

The inductive reactance is proportional to the frequency while capacitive reactance decreases as frequency increases.

Figure 1.38 (b) shows the variation of frequency with current a maximum value at the resonant frequency.

The maximum $\left(I_{max}\right)$ current for RLC circuit, under this condition, become:

$$I_{max} = \frac{V_s}{R} \quad (1.74)$$

The relative damping in an oscillator is expressed by a constant Q, known as quality factor.

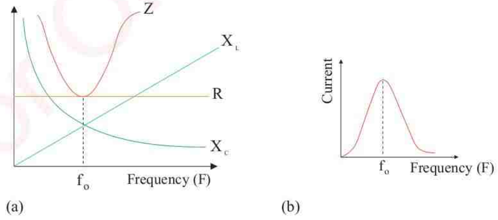


Figure 1.38 Resonance characteristics curve

Thus, the less the damping the larger the Q value and vice versa. For an oscillator with resonant angular frequency ω_o, Q is dimensionless and given as a ratio:

$$\text{Q factor} = \omega_o \frac{\text{energy stored}}{\text{Average power dissipated}}$$

$$Q = \omega_o \frac{L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}} \qquad (1.75)$$

The Q factor describes the width of the current peak shown in Figure13.9. According to this figure, the bandwidth ($\Delta\omega$) of the resonance peak is well-defined as the range of angular frequencies ω over which the average power P_{av} is greater than one-half the maximum value of P_{av}.

Similarly the quality factor Q of the circuit can be defined as; $Q = \dfrac{\omega_o}{\Delta\omega}$

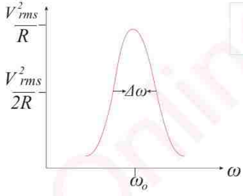


Figure 1.39 RLC circuit peaks at the resonant angular frequency

(d) Power in RLC circuits

In a circuit in which there is resistance, inductance, and capacitance the total average power consumed (P_{av}) is equal to that dissipated in the resistor only. Therefore expression for dissipated power in *a.c.* circuits is given by:

$$P_{av} = I^2_{rms} R$$

Power factor

Power is only consumed by the resistive part of the circuit that is in the expression *IV* where *V* is applied *p.d.* across the total resistance in the circuit and I is the instantaneous current. If E_{rms} is the *p.d.* applied to an *a.c.* circuit leading to the current I_{rms} by phase angle ϕ, the instantaneous electric power in the circuit is given by:

$$P_i = EI = E_o I_o \sin\omega t \sin(\omega t + \phi)$$

where $I = I_o \sin(\omega t + \phi)$ and $E = E_o \sin\omega t$

Using the trigonometric identity

$$\sin(\omega t + \phi) = \sin\omega t\cos\phi + \sin\phi\cos\omega t$$

The power becomes

$$P_i = E_o I_o \sin^2\omega t\cos\phi + E_o I_o \sin\omega t\sin\phi\cos\omega t$$

The average over a periodic time of the second term on the right hand side of the equation above is zero. The average of the first one term of the right side of the equation above is

$$P_{av} = E_o I_o \cos\phi \frac{\int_0^T \sin^2\omega t\, dt}{T}$$

Making use of the trigonometric identity, $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$ it follows that:

$$\frac{\int_0^{2\pi} \sin^2\omega t\, dt}{2\pi} = \frac{\int_0^{2\pi} \frac{1}{2}(1 - \cos 2\omega t)\, dt}{2\pi}$$

$$= \frac{1}{2}\frac{\left\{\left[t\right]_0^{2\pi} - \left[\frac{1}{2\omega}\sin 2\omega t\right]_0^{2\pi}\right\}}{2\pi}$$

$$= \frac{1}{2}\left\{\frac{(2\pi - 0)}{2\pi}\right\} = \frac{1}{2}$$

Hence,

$$P_{av} = \frac{1}{2}E_o I_o \cos\phi = \frac{E_o}{\sqrt{2}}\frac{I_o}{\sqrt{2}}\cos\phi$$

$$= E_{rms} I_{rms} \cos\phi \qquad (1.76)$$

Equation (1.76) is true for the average power in one complete cycle which also represents the average power delivered. The term cos ϕ is called the power factor of the circuit.

Generally,

$$\text{Power factor} = \frac{\text{Real power}}{\textit{Apparent power}}$$

$$= \frac{E_{rms} I_{rms} \cos\phi}{E_{rms} I_{rms}} = \cos\phi$$

Example 1.24

Consider an LCR series circuit with R= 300 Ω, L = 0.9 H, C = 2.0 μF where the source frequency is 50 Hz and the voltage is 240 V. Determine:

(a) the circuit impedance;

(b) the current through the circuit;

(c) the phase angle; and

(d) the voltage across each component.

Solution

(a) $Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)}$

$$Z = \sqrt{90000\,\Omega^2 + (282.7\,\Omega - 1591.5\,\Omega)^2}$$

$$= 1342.7\ \Omega$$

$$Z = 1342.7\ \Omega$$

(b) $I = \frac{E}{Z} = \frac{240.0\text{ V}}{1342.7\Omega} = 0.18\text{ A}$

$$I = 0.18\text{ A}$$

(c) $\phi = \tan^{-1}\left(\frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}\right)$

$$= \tan^{-1}\left(\frac{(282.7\,\Omega - 1591.5\,\Omega)}{300\,\Omega}\right) = -77\,°$$

$$\phi = -77\ \Upsilon$$

Since ϕ is negative, the circuit is capacitive.

(d) $V_R = IR = 0.18\text{A} \times 300\Omega = 54.00\text{ V}$

$$V_L = IX_L = 0.18\text{A} \times 282.7\,\Omega$$
$$= 50.89\text{ V}$$

$$V_C = IX_C = 0.18\text{A} \times 1591.5\,\Omega$$
$$= 286.47\text{ V}$$

Example 1.25

(a) What is the resonant frequency of RLC circuit with resistance 5Ω, inductance 3 mH and capacitance 1 μF ?

(b) If an *a.c.* source of constant voltage amplitude 4 V is set to this frequency, what is the average power transferred to the circuit?

(c) Determine the *Q* factor of this circuit.

(d) Determine the bandwidth of this circuit.

Solution

(a) The resonant frequency is

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{3\times10^{-3}\text{H}\times1\times10^{-6}\text{F}}}$$

$$f = 2.905\,\text{kHz}$$

(b) The average power transferred to the circuit is a maximum therefore

$$P_{av} = \frac{V_{rms}^2}{R} = \frac{\left(\frac{V_\circ}{\sqrt{2}}\right)^2}{R} = \frac{(4\,\text{V})}{2\times5\Omega} = 1.6\,\text{W}$$

$$P_{av} = 1.6\text{ W}$$

(c) The Q factor

$$Q = \frac{L\omega_o}{R} = \frac{3\times10^{-3}\text{H}\times2\pi\times2905\,\text{Hz}}{5\Omega}$$

$$=10.95$$

$$Q = 10.95$$

(d) The bandwidth

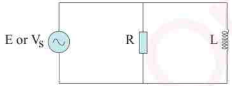
$$\Delta\omega = \frac{\omega_\circ}{Q} = \frac{2\pi\times2905\,Hz}{10.95} = 1666.91\text{ rad/s}$$

$$\Delta\omega = 1666.91\,\text{rad/s}$$

(e) Inductor-Resistor (LR) circuit in parallel

Figure 1.40 shows an LR parallel circuit connected to an *a.c.* power supply. As the elements are in parallel, the current through resistor is I_R and current through inductor is I_L. Therefore the current in all components has different phase and amplitude as indicated in Figure 1.40(b); the current through the inductor is $\pi/2$ out of phase with the current through the resistor, I_R, as I_R is in phase with V, then I_L lags the voltage V by $\pi/2$.

(a)



(b)

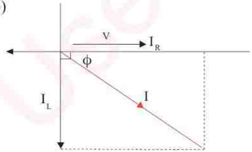


Figure 1.40 LR parallel circuit and phasor diagram

For a RL parallel circuit we add the currents by the vector method taking into account the phase angle between them. Thus,

$$I^2 = I_L^2 + I_R^2 = \left(\frac{V}{X_L}\right)^2 + \left(\frac{V}{R}\right)^2$$

$$I = V\sqrt{\frac{1}{(\omega L)^2} + \frac{1}{R^2}} \qquad (1.77)$$

The impedance for LR circuit in parallel is given as:

$$Z = \frac{V}{I} = \frac{1}{\sqrt{\frac{1}{(\omega L)^2} + \frac{1}{R^2}}} \qquad (1.78)$$

Consider ϕ as the phase angle between I and V for the parallel LR circuit since I, lags the applied voltage V by an angle ϕ then:

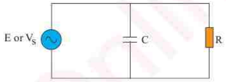
$$\tan\phi = \frac{I_L}{I_R} = \frac{V}{X_L} \times \frac{R}{V} = \frac{R}{X_L} = \frac{R}{\omega L}$$

$$\phi = \tan^{-1}\left(\frac{R}{\omega L}\right) \qquad (1.79)$$

(f) Capacitor Resistor (CR) circuit in parallel

Figure 1.41(a) shows a CR parallel circuit. The supplied current is the vector sum of the current through the resistor, I_R and current through the capacitor, I_C. The current through the resistor is $\pi/2$ out of phase with current through the capacitor. Since I_R is in phase with V, then I_C leads the applied voltage V by $\pi/2$ Figure 1.41(b).

(a)



(b)

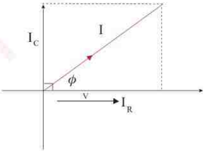


Figure 1.41 CR in parallel circuit and its phasor diagram

The vector sum of currents is

$$I^2 = I_C^2 + I_R^2 = \left(\frac{V}{X_C}\right)^2 + \left(\frac{V}{R}\right)^2$$

$$I = V\sqrt{(\omega C)^2 + \frac{1}{R^2}} \qquad (1.80)$$

The impedance for CR circuit in parallel is given as:

$$Z = \frac{V}{I} = \frac{1}{\sqrt{(\omega C)^2 + \frac{1}{R^2}}} \qquad (1.81)$$

Let angle ϕ be the phase angle between I and V for a parallel CR circuit. The current I leads the applied voltage V by an angle ϕ, thus:

$$\tan\phi = \frac{I_C}{I_R} = \frac{V}{X_C} \times \frac{R}{V} = \frac{R}{X_C} = \omega CR$$

$$\phi = \tan^{-1}(\omega CR) \qquad (1.82)$$

(g) Inductor Capacitor (LC) circuit in parallel

Consider LC circuit connected in parallel with *a.c.* power supply, shown in Figure 1.42(a). The instantaneous *p.d.* applied across each element is the same, vector V. Figure 1.42(b) shows that, I_C leads V by $\pi/2$ while I_L lags V by $\pi/2$.

For a net capacitive circuit i.e. I_C greater than I_L the net current is given as

$$I = I_C - I_L = \frac{V}{X_C} - \frac{V}{X_L}$$

$$I = V\left(\frac{1}{X_C} - \frac{1}{X_L}\right) = V\left(\omega C - \frac{1}{\omega L}\right)$$

$$I = V\left(\omega C - \frac{1}{\omega L}\right) \qquad (1.83)$$

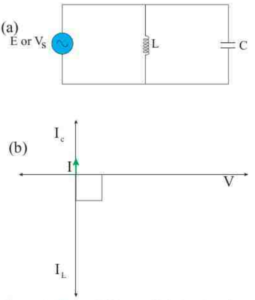


Figure 1. 42 LC in parallel circuit and its phasor diagram

The impedance for the LC circuit in parallel is given as

$$Z = \frac{V}{I} = \frac{1}{\left(\omega C - \frac{1}{\omega L}\right)}$$

$$Z = \left(\omega C - \frac{1}{\omega L}\right)^{-1} \qquad (1.84)$$

Similarly, the net inductive circuit that is (I_L) greaterthan (I_C) the net current (*I*) and impedance (Z) for the LC circuit can be expressed respectively as:

$$I = V\left(\frac{1}{\omega L} - \omega C\right)$$

$$Z = \frac{V}{I} = \left(\frac{1}{\omega L} - \omega C\right)^{-1} \qquad (1.85)$$

Figure 1.43(a) shows how the current along the capacitor and inductor varies with the frequency. The magnitude and phase of current vary according to the relative magnitude of capacitive and reactive reactance (which depends on frequency) as shown in Figure 1.43 (b), while in Figure 1.43 (c) illustrates how the impedance of the parallel L and C circuits varies with frequency.

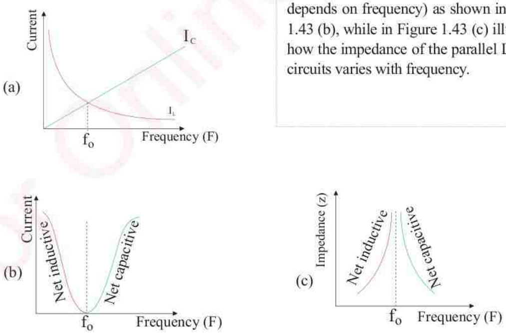
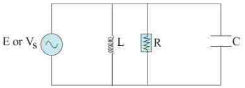


Figure 1.43 Variation of current and impedance for LC in parallel circuits

(h) Resistor, Inductor and Capacitor (RLC) circuit in parallel

Consider RLC circuit connected in parallel with an *a.c.* power supply, shown in Figure 1.44 (a). The instantaneous applied *p.d.* across each element is the same, vector 'V', the terminal alternating voltage. Figure 1.44 (b) shows that, I_R is in phase with V, I_L is lagging V by $\pi/2$ and I_C leading V by $\pi/2$.

(a)



(b)

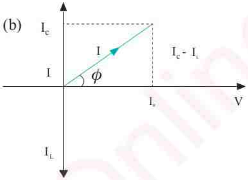


Figure 1. 44 RLC in parallel circuit and its phasor diagram

The net current I is equal to the vector sum of instantaneous currents and is written as

$$I^2 = I_R^2 + \left(I_C - I_L\right)^2$$

$$I = V\sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}$$

The impedance for RLC circuit in parallel is given as:

$$Z = \frac{V}{I} = \frac{1}{\sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}}$$

$$= \left(\sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}\right)^{-1} \quad (1.86)$$

Since resonance, $\omega C = 1/\omega L$, the impedance for RLC circuit in parallel is net resistive.

Therefore, $Z = R$

Also at resonance, the net current I in RLC parallel circuit is minimum as compared to RLC in series which has maximum current. In parallel circuit the currents in L and C are always exactly a half cycle out of phase, similarly they also have equal magnitudes and cancel each other.

The total current is purely that through R thus $I = \frac{V}{R}$. It is known that $X_C = X_L$ and therefore cancel at resonance, this does not mean there is no current through the capacitor or inductor. Assume the value of R is large, then the equivalent impedance of the circuit near resonance is much larger than the individual inductive reactance and capacitive reactance.

Example 1.26

A person is made to walk through the entrance of a metal detector, for security reasons. If the person is carrying anything made of a metal, the metal detector emits a sound waves. Explain the principle of operation of the metal detector.

Solution

The metal detector works on the principle of resonance in *a.c.* circuits. The system is made of a coil of many turns. The coil is connected to a capacitor tuned so that the circuit is at resonance. When one walks through the ditector with a metal in the pocket, the impedance of the curcuit changes resulting to a signficant change in the current flowing through the curcuit. This change in current is detected and the electronic circuitry tiggers an alarm.

Exercise 1.3

1. Calculate the impedance of a 100 mH inductor of negligible resistance at frequency:
   a) 10 Hz;
   b) 50 Hz; and
   c) 1 MHz.
2. Distinguish between the reactance and the resistance of a coil, and show how they are related.
3. An 8 μF capacitor is placed across a 200 V, 50 Hz supply. Determine:
   a) the root mean square current that flows in the circuit; and
   b) the peak value of the voltage across the capacitor.
4. A coil of inductance 0.2 mH and negligible resistance is connected in parallel with a capacitor of capacitance 0.005 μF. Calculate the impedance if angular frequency of the *a.c.* source is 20 rad/s.
5. A 4 mH inductor, 5 ∝F capacitor and 54 Ω resistor are connected in parallel with a 50 V, 10 kHz *a.c.* source. Find the value of the impedance and current in the circuit.
6. A 44 mH inductor is connected to 240 V, 50 Hz *a.c.* supply.
   a) Find the rms of the current in the circuit.
   b) Determine the net power absorbed over a complete cycle. Explain your answer.
7. A 60 ∝F capacitor is connected to a 240 V, 60 Hz *a.c.* supply.
   a) Calculate the rms value of the current in the circuit.
   b) Determine the net power absorbed over a complete cycle. Explain your answer.
8. A series LCR circuit with $R = 20\,\Omega$, $L = 1.5\,H$ and 35 ∝F is connected to a variable frequency 200 V *a.c.* supply. When the frequency of the supply is equal to the natural frequency of the circuit, what is the average power transferred to the circuit in one complete cycle?
9. A coil of inductance 0.50 H and resistor of a resistance 100 Ω are connected in series to a 240 V, 50 Hz *a.c.* supply.
   a) What is the maximum current in the coil?
   b) What is the time lag between the maximum voltage and maximum current?

10. Obtain the resonance frequency of a series LCR circuit with $C = 30\mu F$, $R = 15\,\Omega$ and $L = 5$ H. What is the Q-value and bandwidth for this circuit?

Revision Exercise

1. What does conduction of electrons mean? Although the drift velocity of electrons is very small, the lamp lights up as soon as a switch is put on. Explain this observation.

2. A potential difference of 9.0 V is causing electrons to flow through a steel wire so that 1.0×10^{20} electrons pass through a point in the wire in 60 s. Calculate:
   (a) The charge which passes the point in 60 s
   (b) The electric current in the wire
   (c) The resistance of the wire.

3. (a) Assume that there is one free electron per atom. Calculate the number of free electrons in a piece of silver of cross-section area $1.5\times10^{-4}\,m^2$ and length 2 m. Atomic weight of silver is 108 and density of silver is $1.05\times10^{4}\,kgm^{-3}$.

   (b) Determine the drift velocity of electrons in a copper conductor having a cross-sectional area of $5\times10^{-6}\,m^2$ if the current is $10\,A$. (Assume that there are 8×10^{28} electrons/m^3).

4. (a) A piece of a wire has a resistance R. What will be the resistance of a wire of the same metal which is three times as long and twice as thick?

   (b) A wire of length 45.5 cm has a resistance of 0.50 Ω. What will be its resistance when it is drawn out to a length of 54.6 cm, assuming that the volume remains constant?

5. Define the temperature coefficient of resistance and explain its significance in current conduction through metals. A certain coil of wire has an electrical resistance of 24 Ω at 10 °C, and at 20 °C the resistance increases to 28 Ω. Compute the temperature coefficient of resistance for the metal of which the coil is made.

6. A copper coil has resistance of 20 Ω at 0 °C and 28 Ω at 100 °C.
   (a) What is the temperature coefficient of resistance of copper?

   (b) The wire is used in a circuit and when an *e.m.f.* of 12 V is connected across it, the power produced is 6 W. What is the temperature of the coil?

7. The heating coil of power rating 10 W is required when the *p.d.* across it is 20 V.
   (a) Calculate the length of a constantan wire needed to make the coil if the cross sectional area of the wire used is $1\times10^{-7}\,m^2$.

(b) What length of wire would be required if its diameter is a halved?

8. (a) Define the electromotive force, and explain the meaning of the internal resistance of a battery.

(b) A high resistance voltmeter reads 1.5 V when connected to the terminals of a battery. When the battery is supplying a current of 0.30 A through an external resistance R, the voltmeter reads 1.2 V. Calculate:

(i) the e.m.f. of the battery;

(ii) the value of R;

(iii) the internal resistance of the battery; and

(iv) the energy converted from chemical to electrical energy by the battery in 2.0 s.

9. The following are four electrical components:

A. Component which obeys Ohm's law;

B. Component which obeys Ohm's law but which has higher resistance than the component A;

C. Filament lamp; and

D. Component, other than a filament lamp, which does not obey Ohm's law.

(i) For each of these components, sketch current - voltage characteristics showing both positive and negative values. Use one set of axes for A and B, and separate sets of axes for C and D. Label your graphs clearly.

(ii) Explain the shape of the characteristic for C.

(iii)Name the component you have chosen for D.

10. (a) Explain the differences between *e.m.f.* and Potential difference.

(b) A 24 V battery of internal resistance 4Ω is connected to a variable resistor. What value of the current in drawn from the battery if the rate of the heat produced in the resistor is maximum.

11. An accumulator of *e.m.f.* 2 V and negligible internal resistance is connected in series with 500 Ω and Y Ω resistors. The reading of the voltmeter across 500 Ω and Y Ω are $\frac{2}{7}$ V and $\frac{8}{7}$ V respectively. Calculate the value of Y and the voltmeter resistance.

12. Figure 1.45 shows a battery of 10 V and negligible internal resistance connected across the diagonally opposite corners of a cubical network consisting of 12 resistors each of resistance 1 Ω. Determine:

(a) the equivalent resistance of the network; and

(b) the total current supplied by the battery.

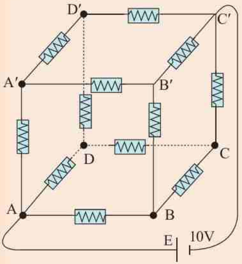


Figure 1.45 A cubical electrical network

13. From the circuit diagram shown in Figure 1.46, find; the values of I_1, I_2, I_3, I_4 and I_5.

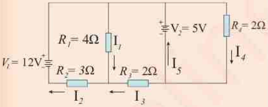


Figure 1.46 An electrical network

14. Calculate V_1, V_2 and R_5, shown in Figure 1.47.

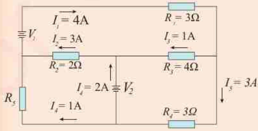


Figure 1.47 An electrical network

15. In the circuit given in Figure 1.48 a battery of e.m.f. 12.6 V and internal resistance 0.1 Ω is being charged from *d.c.* source of *e.m.f.* 24 V and internal resistance 1.0 Ω. V_1 and V_2 are high resistance voltmeters and *R* is a fixed resistor.
    (a) If the charging current is 5.0 *A*, find the resistance of the resistor *R*.
    (b) If the resistance of the resistor *R* were changed to 0.9 Ω, what would be the reading of each voltmeter?

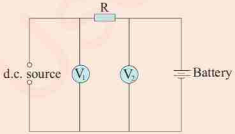


Figure 1.48 A sketch of electrical circuit

16. (a) Describe four uses of a potentiometer.
    (b) Give three advantages and disadvantages of a potentiometer.
    (c) Explain the meaning of (e.m.f.).
    (d) The e.m.f. of a battery A is balanced by length of 75.0 cm on a potentiometer wire. The *e.m.f.* of a standard cell 1.02 *V* is balanced by a length of 50.0 cm.
    (i) What is the *e.m.f.* of A?
    (ii) Calculate the new balance length if A has internal resistance of 2 Ω and a resistor of 8 Ω is joined to its terminals.

17. (a) A simple potentiometer circuit is set up as shown Figure 1.49, using a uniform wire MN 1.0 m long and resistance 2.0 Ω. Assume the internal resistance of 4 V battery is negligible and the variable resistor R = 2.4 Ω. What would be the length on MN for zero galvanometer deflection?

(b) Assuming R = 1.0 Ω in Figure 1.49 and voltmeter of 2 Ω replace 1.5 V cell and the galvanometer. What would be the reading of the voltmeter, if the contact L were placed at the mid-point of MN?

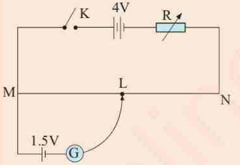


Figure 1.49 A simple potentiometer circuit

18. Consider a Wheatstone bridge circuit shown in Figure 1.50, the four resistors in the arms of the bridge are $R_1 = 10\ \Omega$, $R_2 = 3\ \Omega$, $R_3 = 2\ \Omega$ and $R_4 = 4\ \Omega$, . If a terminal cell E = 2 V with negligible internal resistance is used in the circuit and a galvanometer of resistance 10.0 Ω is connected between C and D. What will be the current in the galvanometer?

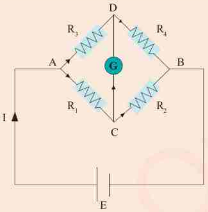


Figure 1. 50 Wheatstone circuit bridge

19. A 2V potentiometer used for determination of internal resistance of a 1.5 V cell is shown in Figure 1.51. The balance point of the cell in open circuit is 76.3 cm. When a resistor 9.5 Ω is used in the external circuit of the cell, the balance point shifts to 64.8 cm length of the potentiometer wire. Find the internal resistance of the cell.

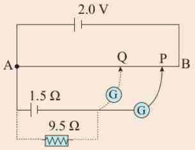


Figure 1. 51 A potentiometer

20. A circuit is setup as shown in Figure 1.52. When $R_1 = 10\ \Omega$ and $R_2 = 90\ \Omega$ there is no current through the galvanometer G. What is the resistance of P, assuming the resistance of the cells are negligible?

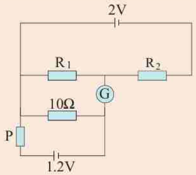


Figure 1.52 An electrical network

21. Explain the importance of ionization agent and chain ionization in self-sustaining gas discharge.
22. Explain the importance of ionization agent and chain ionization in initiated gas discharge. Can the current be constant as voltage rises?
23. Explain the term saturated current. Does it need an ionization agent? Explain your answer.
24. Differentiate between direct current and alternating current. An *a.c.* voltage supplied to our homes is given by $E = E_o \sin \omega t$, but we say that *a.c.* voltage in our home is 240 V, 50 Hz. Explain.
25. Why a *d.c.* ammeter cannot read *a.c.*? Give the physical explanation for the phase difference between voltage and current in inductive and capacitive circuits.
26. An *a.c.* series circuit consists of a capacitor of 2 μF and a resistor of 1000 Ω. An alternating e.m.f. of 12 V and frequency 50 Hz is applied. Find:
    (a) the current flowing through the circuit;
    (b) the voltage across the capacitor;
    (c) the phase angle between the applied *e.m.f.* and the current;
    (d) the average power developed in the circuit.
27. Calculate the resonance frequency and determine the Q-value of a series LRC circuit with $R = 10\ \Omega$, $L = 2\ H$ and $C = 32\ \propto$F.
28. Determine the phase angle and power factor for a RLC series circuit containing a resistor of 50 Ω, a capacitor of $10\ \propto$F and an inductor of 0.45 H, when connected to a power supply with frequency of 60 Hz.
29. A coil carries a steady current of 50 mA when a steady *d.c.* voltage across it is 2.0 V. The steady e.m.f. is replaced by an *a.c.* of voltage 4.1 V, and a frequency of 50 Hz. The current flowing through the coil is found to be 11.0 mA. Calculate:
    (a) the resistance of the coil;
    (b) the self-inductance of the coil; and
    (c) the phase difference between the alternating current and the alternating (e.m.f.).
30. A capacitor of capacitance *C*, a coil of inductance *L*, resistor of resistance *R*, and a lamp are placed in parallel with an alternating voltage V. Its frequency *f* is varied from zero to a high value while the magnitude of V is kept constant. Describe how the brightness of the lamp would vary.

Chapter Two

Electromagnetism

Introduction

The concept of electromagnetism is a combination of two concepts, namely electricity and magnetism. In daily life, people interact with several devices that make use of electromagnetism. Such devices include generators, computer hard disks, music cassettes, video tapes, digital bank cards, automatic vending machines, and many others. In this chapter, you will be introduced to the basic principles of operation of these devices by learning magnetic fields and sources, forces of magnetic fields, the relationship between magnetic fields and electric currents, as well as magnetic properties of materials. You will also learn about the magnetic fields of the earth.

2.1 Magnetic fields and forces

A magnetic field is defined as a region where a moving charge or a current carrying conductor experiences a force. Given the direction of the magnetic field lines, the direction of the magnetic force on a moving charge can be determined using the Fleming's left-hand rule. For a negatively charged particle the direction of its velocity is opposite to the direction of conventional current. Knowledge of the direction of velocity of the moving charge and the magnetic flux density makes it possible to compute the magnitude of this force.

2.1.1 Magnetic force

In electrostatics, the interaction between charges is described in terms of electric intensity $\vec{E}$ in which a charge q at rest in the field experiences a force $\vec{F} = q\vec{E}$ in the direction of the electric field. In magnetism, permanent magnet or moving charges are sources of magnetic field. When a charge traverses through a magnetic field, it experiences a force $\vec{F}$ whose magnitude and direction depends on the magnitude of the charge q, the magnetic flux density $\vec{B}$ and the velocity $\vec{v}$ of the charge. The force is given by (Lorentz force):

$$\vec{F} = q\vec{v} \times \vec{B} \qquad (2.1)$$

From this equation, the magnetic flux density at some point in space can be defined in terms of the force exerted on a charged particle moving with velocity $\vec{v}$. The magnetic flux density is a measure of concentration of magnetic lines per unit area and therefore it is an appropriate quantity to describe the magnetic field in space. The SI unit of the magnetic flux density is Tesla (T) which can be deduced from the following equation.

$$B(T) = \frac{F(N)}{q(As)v(m/s)} \qquad (2.2)$$

From this equation, one Tesla can be defined as the flux density in the field when a force of 1 N is acting on a conductor, of 1 m long, placed perpendicular to the field and carrying a current of 1 A.

As shown in Figure 2.1, the force $\vec{F}_B$ on a moving charge is perpendicular to both $\vec{v}$ and $\vec{B}$.

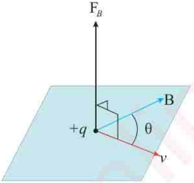


Figure 2.1 Relationship between the magnetic force velocity and magnetic flux density.

Using the Cartesian coordinates with $v\hat{i}$ and $B\hat{j}$, the direction of the force can be deduced from the cross product (with $\theta = 90^0$):

$$\vec{F} = q\left(v\hat{i} \times B\hat{j}\right) = qvB\left(\hat{i} \times \hat{j}\right) = qvB\hat{k} \qquad (2.3)$$

Remember the order of cross multiplication of the unit vectors that produces positive unit is anticlockwise as illustrated in Figure 2.2.

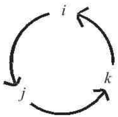


Figure 2.2 Order of cross product of unit vectors

The direction of the force defined by the unit vector $\hat{k}$ in the $z-$direction can also be determined using the left hand rule illustrated in Figure 2.3. In this rule, the direction of the first finger points in the direction of the field, the direction of the second finger points in the direction of the conventional current and the thumb points in the direction of the force (thrust).

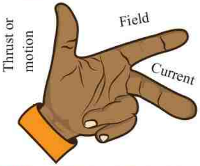


Figure 2.3 Illustration of the Fleming's left hand rule.

Activity 2.1

Identifying the direction of $\vec{F}$ in the following cases given the directions of $\vec{v}$ and $\vec{B}$, using the left-hand rule.

1. The magnetic field is pointing towards the board.
2. The negatively charged particle is moving from left to the right of the board.

3. The positively charged particle is moving from left to right of the board.

Note: If the sign of the charge changes from positive to negative, the direction of the force is reversed.

By definition of a cross product given in equation (2.3), the magnitude of the force is given by

$$\left|\vec{F}\right| = q\left|\vec{v}\right|\left|\vec{B}\right|\sin\theta \qquad (2.4)$$

where θ is the angle between $\vec{v}$ and $\vec{B}$. From this equation, it is evident that if the velocity of the charged particle is parallel $\left(\theta = 0^\circ\right)$ to the magnetic flux density, the magnetic force on the charged particle will be zero. Maximum magnetic force is experienced when the velocity is perpendicular $\left(\theta = 90^\circ\right)$ to the magnetic flux density.

It is important to note that the direction of the force is dependent on the sign of the charge as indicated in Figure 2.4. This is expected since replacement of $+q$ by $-q$ in Figure 2.1 would reverse the direction of force in equation (2.3).

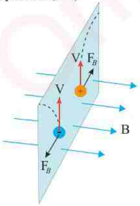


Figure 2.4 The magnetic force on a moving charge in $\vec{B}$

Example 2.1

A particle with a charge of 3 μC is moving with a speed of 5×10^6 m/s perpendicular to a magnetic field of strength 0.08 T. Determine the magnitude of the force on the particle.

Solution

The force on the particle is given as

$\vec{F} = q\vec{v}\times\vec{B}$

The magnitude of the force is

$F = qvB\sin\theta$

$F = \left(3\times10^{-6}\text{ C}\right)\left(5\times10^{6}\text{ ms}^{-1}\right)\left(0.08\text{ T}\right)\times\sin 90^\circ$

$= 1.2\text{ N}$

Example 2.2

A proton moving at 2.8×10^6 m/s through a magnetic field of *170* T experiences a magnetic force of magnitude 7.3×10^{-13} N. Calculate the angle between the proton's velocity and the field.

Solution

$F = qvB\sin\theta \Rightarrow \theta = \sin^{-1}\left(\dfrac{F}{qvB}\right)$

$\theta = \sin^{-1}\left(\dfrac{7.3\times10^{-13}\text{ N}}{1.6\times10^{-19}\text{ C}\times2.8\times10^{6}\text{ ms}^{-1}\times1.70\text{ T}}\right)$

$= 73.4^\circ$

Example 2.3

An electron that has velocity

$$\vec{v} = \left(3\times10^{6}\ \text{m/s}\right)\hat{\text{i}} + \left(2\times10^{6}\ \text{m/s}\right)\hat{\text{j}}$$

moves through the magnetic field .

$$\vec{B} = \left(0.17\,T\right)\hat{i} + \left(0.1\,T\right)\hat{j}$$

(a) Find the force on the electron

(b) What will be the force if the motion was for a proton instead of an electron?

Solution

(a) $\vec{\text{F}} = \text{q}\vec{\text{v}}\times\vec{\text{B}}$ (Note that $\text{q} = -\text{e}$)

$$\vec{\text{F}} = -1.6\times10^{-19}\left(3\times10^{6}\hat{i} + 2\times10^{6}\hat{j}\right)\times\left(0.17\hat{i} + 0.1\hat{j}\right)$$

$$\vec{\text{F}} = -1.6\times10^{-19}\left(\left(3\times10^{6}\times0.1\,\hat{k}\right) + \left(2\times10^{6}\times0.17\left(-\hat{k}\right)\right)\right)$$

$$\vec{\text{F}} = -1.6\times10^{-19}\left(-4\times10^{4}\,\hat{k}\right) = \left(6.4\times10^{-15}\hat{k}\right)N$$

(b) The same as in (a) but in the $-\hat{k}$ direction.

2.1.2 Motion of a charged particle in a uniform magnetic field

As seen in the previous section, a charged particle experiences a force when it moves through a magnetic field. The question that was not addressed in the previous section is; what happens to a charged particle which enters a uniform magnetic field with velocity at different inclinations to the field? To answer this question, several scenarios will be considered. The simplest case occurs when a charged particle enters a uniform magnetic field at right angle. Assuming the motion of the charged particle takes place in vacuum, the magnetic field is the dominant factor determining the motion. The magnetic force is always perpendicular to velocity and hence no work is done by the magnetic force on the charged particle. Using the work energy theorem, it is evident that the particle's kinetic energy is constant and consequently only the direction of motion is affected. The force therefore tends to deflect the particle in a curved path. If the magnetic field is constant and since the speed is constant the deflection is also constant and consequently the particle continues to follow this circular path as shown in Figure 2.5.

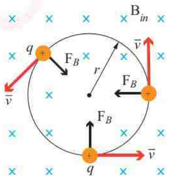


Figure 2.5 Circular path traversed by a charged particle in a constant magnetic field B_{in}

Since the charged particle has mass m, it will experience a centripetal force given by:

$$F_c = \frac{mv^2}{r} \tag{2.5}$$

According to equation (2.4), the magnitude of the magnetic force in Figure 2.5 becomes:

$$F = qvB \tag{2.6}$$

The centripetal force is supplied by the magnetic force, F hence, at the equilibrium

$$qvB = \frac{mv^2}{r} \text{ or } r = \frac{mv}{qB} \qquad (2.7)$$

It is evident from equation (2.7) that the radius of the circular path in Figure 2.5 is proportional to the linear momentum *(mv)* of the charged particle and inversely proportional to both the magnitude of its charge (q) and magnetic flux density, $\vec{B}$. The angular speed, ω of the particle can be deduced from equation (2.7) and written as

$$\omega = \frac{v}{r} = \frac{qB}{m} \qquad (2.8)$$

The angular speed ω is related to the period of the circular motion of the charged particle as follows:

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{qB} \qquad (2.9)$$

From equations (2.8) and (2.9), it is clear that the angular speed of the charged particle and the period of its circular motion are independent of both the radius of the orbit and the linear speed. This concept has useful applications in various areas including the cyclotron accelerator.

Consider a case where the charged particle enters a uniform magnetic field with velocity v at some arbitrary angle θ Figure 2.6.

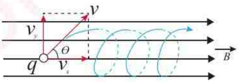


Figure 2.6 A charged particle forms a helical trajectory.

In this situation the velocity can be resolved into x component, $v_x = v\cos\theta$ in the direction of the magnetic field and y component $v_y = v\sin\theta$ perpendicular to the direction of the magnetic field. The effects of these components on the motion of the particle are quite different. The y component moves the charged particle in a circular path as previously described while the x component moves the charged particle in the direction of the magnetic field. Consequently the charged particle forms a ***helical path*** along the direction of the magnetic field as seen in Figure 2.6.

It is useful to observe that the perpendicular component v_y determines the quantities that influence circular motion while the horizontal component v_x determines the linear distance covered by the charged particle after completing one cycle also known as the pitch of the helix. From equation (2.7), the radius of the path is

$$r = \frac{mv}{qB}\sin\theta \qquad (2.10)$$

Similarly, from equation (2.9), the frequency f as a reciprocal of the period is given by:

$$f = \frac{qB}{2\pi m} \text{ or } \omega = \frac{qB}{m} \qquad (2.11)$$

The pitch (p) of the helix as the linear distance covered by the charged particle upon completion of one circular revolution or the linear distance covered by the charged particle during the periodic time T is given by;

$$p = (v\cos\theta)\left(\frac{2\pi m}{qB}\right) \qquad (2.12)$$

Example 2.4

A proton moves at 8.5×10^{7} m/s perpendicular to a magnetic field. The fields causes the proton to travel in a circular path of radius 0.68 m Determine the magnetic field strength. (Mass of proton $m_p = 1.67\times10^{-27}$ kg and its charge, $q = 1.6\times10^{-19} C$).

Solution

For a proton to travel in a circular path, the centripetal force has to be supplied by the magnetic fields and hence, at equilibrium

$$qvB = \frac{mv^2}{r} \Rightarrow B = \frac{mv}{qr}$$

$$B = \frac{1.67\times10^{-27}\,\text{kg}\times8.5\times10^{7}\,\text{ms}^{-1}}{1.6\times10^{-19}\,\text{C}\times0.68\,\text{m}}$$

$$=1.3\,\text{T}$$

Example 2.5

An electron moving along the $+x$ axis at a speed $v=5.5\times10^{6}\ m/s$ enters a uniform magnetic field that makes an angle of 72° with the $+x$ axis as shown in Figure 2.7. If the magnetic field of magnitude B has the magnitude of 0.32 T, calculate:

(a) the pitch p;
(b) the radius R of the trajectory; and
(c) the time required for one trip around the helix.

(Mass of electron, $m_e = 9.11\times10^{-31} kg$ and electron charge, $q = 1.6\times10^{-19} C$)

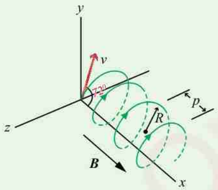


Figure 2.7 An electron moving in a helical path of radius R.

Solution

(a) The pitch or distance between loops p is given by

$$p = v\cos\theta\left(\frac{2\pi m}{qB}\right)$$

$$p = \frac{2\pi\times9.11\times10^{-31}\,\text{kg}\times5.5\times10^{6}\,\text{ms}^{-1}\times\cos72^{0}}{1.6\times10^{-19}\,\text{C}\times0.32\,\text{T}}$$

$$=1.9\times10^{-4}\,\text{m}$$

(b) The radius of the trajectory R is given as

$$R = \frac{mv\sin\theta}{qB}$$

$$R = \frac{9.11\times10^{-31}\,\text{kg}\times5.5\times10^{6}\,\text{ms}^{-1}\sin72^{0}}{1.6\times10^{-19}\,\text{C}\times0.32\,\text{T}}$$

$$=9.31\times10^{-5}\,\text{m}$$

(c) The period of the charged particle going around a circle is calculated using a given mass, charge and the magnetic field intensity.

Using the equation for frequency $f = \dfrac{qB}{2\pi m}$ gives;

$$T = \frac{1}{f} \Rightarrow T = \frac{2\pi m}{qB}$$

$$T = \frac{2\pi \times 9.11 \times 10^{-31} \text{kg}}{1.6 \times 10^{19} \text{C} \times 0.32 \text{T}}$$

$$= 1.12 \times 10^{-10} \text{ seconds}$$

2.1.3 Magnetic force on a current-carrying conductor

A current-carrying conductor is essentially charges moving through the conductor. As discussed in the previous section, charges moving in a magnetic field experience a force and therefore current-carrying conductor will also experience the same. It is worth noting that a current-carrying conductor generates a magnetic field which can also exert a force on another current-carrying conductor.

(a) Oersted's experiment

The first evidence of the relationship between magnetic field and a current carrying conductor was discovered in 1820 by the Danish scientist Hans Christian Oersted. He observed that a compass needle was deflected by a current carrying straight wire. From this observation, it was established that electric currents produce magnetic fields. Figure 2.8 (a) showed the compass needle around a wire carrying no current pointed in the same direction toward the earth's north pole. The compass needles was deflected tangentially to form a circle around the wire when current was passed as shown in Figure 2.8(b). These observations imply that magnetic field due to a wire carrying a current forms circular loops.

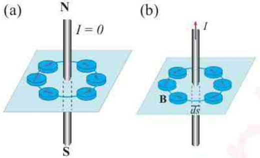


Figure 2.8 Direction of compass needles around a wire

Compass needles Figure 2.8 (a) around a wire points in the same direction (toward the earth's north pole) when no current is passing and Figure 2.8 (b) deflected tangentially to form a circle around the wire when current is passing. To determine the direction of the magnetic field generated by a current-carrying wire the right-hand grip rule is used. The rule states that "If the current carrying conductor is held with the right hand such that the thumb points in the direction of the current, then the curled fingers will indicate the direction of the magnetic field" as depicted in Figure 2.9 (a).

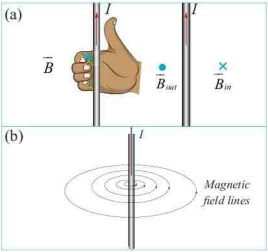


Figure 2.9 Direction of current and magnetic field by right-hand grip rule

By convention, a dot is used to indicate the field lines emerging out of the paper and the cross sign signify field lines going into the paper. Figure 2.9(b) shows a concentric magnetic field lines around a straight wire carrying a current.

(b) Derivation of magnetic force

Consider a wire of length l carrying a current I in a uniform magnetic field $\vec{B}$ and an infinitesimal element of the wire dl as shown in Figure 2.10.

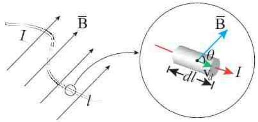


Figure 2.10 A current-carrying wire in a magnetic field

The volume of the infinitesimal element of the wire is $dV = Adl$ where A and dl are the cross-sectional area and the length of the wire element, respectively. The charge contained in this wire element is $dq = (nAdl)e$. From equation (2.1), the force on the charged element is

$$d\vec{F} = (nAdl)e\vec{v}_d \times \vec{B} \qquad (2.13)$$

where n and v_d are the number density (number of charges per unit volume) and drift velocity of the charge carriers respectively. Since the direction of the drift velocity $\vec{v}_d$ is along dl, then, equation (2.13) can be written as

$$d\vec{F} = (nAev_d)\overline{dl} \times \vec{B} \qquad (2.14)$$

The quantity in the bracket is the current I flowing through the wire, hence the magnetic force on the wire segment $\overline{dl}$ is:

$$d\vec{F} = I\overline{dl} \times \vec{B} \qquad (2.15)$$

Equation (2.15) is used to calculate the magnetic force on current-carrying conductors of different configurations. For a straight wire segment in a uniform magnetic field $\vec{B}$, the force on the wire is given by integration of equation (2.15) to get:

$$d\vec{F} = Id\vec{l} \times \vec{B}$$

$$\left|d\vec{F}\right| = I\left|d\vec{l}\right| \times \left|\vec{B}\right|$$

$$dF = IBdl\sin\theta$$

$$\int_0^F dF = IB\sin\theta\int_0^l dl$$

$$F = IBl\sin\theta \qquad (2.16)$$

Where $\vec{l}$ is a vector in the direction of the current I and of magnitude equal to the length l of the wire segment.

Magnitude of magnetic force,

$$\left|\vec{F}\right| = IlB\sin\theta \qquad (2.17)$$

Where θ is the angle between length, l of the wire segment and the magnetic field line of force.

Example 2.6

A straight wire of length 5.6 cm, carrying a current of 36 A passes between the poles of a strong magnet perpendicular to the magnetic of the field and experiences a force of 3.67 N. Calculate the magnetic field strength.

Solution

The magnitude of force experienced is given as $F = IlB \Rightarrow B = \dfrac{F}{Il}$

$$B = \frac{3.67\,\text{N}}{36\,\text{A} \times 5.6 \times 10^{-2}\,\text{m}^{-1}} = 1.82\ \text{T}$$

Example 2.7

A wire of length 16.8 m carries a current of 7 A in a region where a uniform magnetic field has a magnitude of 0.413 T. If the angle between the magnetic field and the current is 62°, as shown in Figure (2.11) what is the magnitude of the magnetic force on the wire?

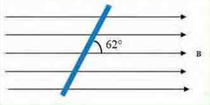


Figure 2.11: A wire in a magnetic field.

Solution

The magnitude of force experienced is given as $F = IlB\sin\theta$

$$\Rightarrow F = 7\text{A} \times 16.8\text{m} \times 0.413\text{T} \times \sin 62$$
$$= 42.9\text{ N}$$

2.1.4 Force and torque on a current loop

A magnetic force on an arbitrarily shaped closed loop is shown in Figure 2.12.

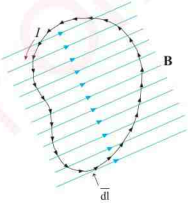


Figure 2.12 Arbitrarily shaped current - carrying loop in a uniform magnetic field

Since this loop is closed the magnetic force acting on the loop is obtained by integrating equation (2.15) over the entire loop. Thus:

$$\bar{F} = I\left(\oint \overline{dl}\right) \times \bar{B} \qquad (2.18)$$

The integral in the bracket is the vector sum of the length elements which forms a closed polygon. Since a vector sum over a closed polygon is zero, the magnetic force $\bar{F}$ on the closed current loop presented in Figure (2.12) is zero. Although the net force in this case is zero, the magnetic field induces a turning effect on closed loops due to a torque which will now be presented.

The turning effect on a current carrying closed loop in a magnetic field is conveniently illustrated using a rectangular coil in a magnetic field at an angle θ to a unit vector normal to the plane of the coil shown in Figure (2.13). Although the net force acting on the loop is zero, the forces on opposite sides of the loop have a turning effect or non-zero net torque because they have different lines of action.

The net torque τ on the current loop shown in Figure 2.13 (a) can be estimated by considering forces F_1 and F_3 along the x-axis, and F_2 and F_4 along the z-axis. The magnetic field B along the y axis is perpendicular to all the forces. Since F_1 and F_3 have the same line of action and equal in magnitude and opposite, the sum of their torques about any axis is $\bar{\tau}_1 + \bar{\tau}_3 = 0$. Force F_2 and F_4 act on the length a of the coil have net torque on the coil because they do not acting on the same line of action.

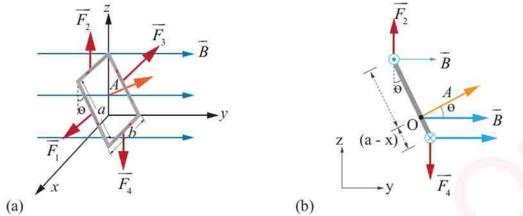


Figure 2.13 (a) A rectangular coil in a uniform magnetic field and (b) its side view

The torque is a product of the force and the perpendicular distance from the axis of rotation passing through point O as shown in Figure 2.13 (b). The distances of F_2 and F_4 from the axis of rotation are x while the corresponding moment of arms are $x\sin\theta$ and $(a-x)\sin\theta$, respectively, hence the net torque on the closed loop is the sum:

$$\sum\bar{\tau} = \bar{\tau}_1 + \bar{\tau}_2 + \bar{\tau}_3 + \bar{\tau}_4$$

$$= F_2 x\sin\theta\bar{i} - F_4(a-x)\sin\theta\bar{i}$$

$$= -IbBx\sin\theta\bar{i} - IbB(a-x)\sin\theta\bar{i} \qquad (2.19)$$

For a number of turns N and the area of the loop $A = ab$, equation (2.19) becomes

$$\tau = -NIAB\sin\theta\bar{i} \qquad (2.20)$$

The product of current and perpendicular area ($I.A$) is known as the magnetic dipole moment, $\propto$ of the loop and therefore equation (2.20) as a vector is given by

$$\bar{\tau} = \bar{\propto}\times\bar{B} \qquad (2.21)$$

The corresponding potential energy (u) is given as $U = -\bar{\propto}\bar{B} = -\propto B\cos\theta_i$

Example 2.8

Find the current through a loop needed to create a maximum torque of 11.8 Nm. The loop has 100 square turns each of length 22 cm and placed in a uniform magnetic field of magnitude 0.96 T.

Solution

For N turns each with area A, the maximum torque is

$\sin\theta i = 1$ then:

$$\tau = NIAB \Rightarrow I = \frac{\tau}{NAB}$$

$$I = \frac{11.8\text{ Nm}}{100\times\left(\frac{22\times10^{-2}\text{m}}{4}\right)^2\times0.96\text{T}} = 40.63\text{ A}$$

Example 2.9

A circular coil of 50 turns and diameter of 0.17 m carrying a current of 5.6 A is suspended vertically in a uniform horizontal magnetic field of magnitude 2 T. The magnetic field makes an angle of 78° with the normal vector of the surface of the coil. Calculate:

(a) The magnitude of the counter torque that must be applied to prevent the coil from turning.

(b) The magnetic dipole moment.

Solution

(a) The counter torque is calculated as

$$\tau = NIAB\sin\theta$$

$$= 50\times 5.6\times \frac{\pi\times(0.17)^2}{4}\times 2\times \sin 78$$

$$= 12.43 \text{ Nm}$$

(b) The counter torque is related to the magnetic dipole moment as

$$\vec{\tau} = \vec{\propto}\times\vec{B} \text{ or } \tau = \left|\vec{\propto}\times\vec{B}\right| = \propto B\sin\theta$$

$$\Rightarrow \propto = \frac{\tau}{B\sin\theta} = \frac{12.43\text{Nm}}{(2\text{T})\sin 78} = 6.35\text{Am}^2$$

2.1.5 Applications of magnetic forces and fields

There are several applications of the principles that a magnetic exerts on a moving charge. These applications are used in some of the devices such as moving coil galvanometer, mass spectrometer and cyclotron, which are described as follows:

(a) Moving coil Galvanometer

The need to measure currents and voltages in circuits arises frequently in daily life. Until recently, the devices used to measure these quantities were exclusively based on the principle of the moving coil galvanometer shown in Figure 2.14.

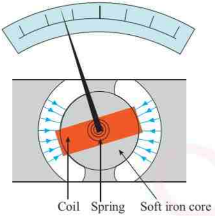


Figure 2.14 Moving coil galvanometer

Nowadays, the functions of these devices are increasingly been replaced by digital device. A moving coil galvanometer consists of a coil with several turns wound on a cylindrical soft iron core, (to strengthen magnetic field generated when a current passes through the coil) mounted on a spring. When a current flows through the coil the magnetic field exerts a torque $\tau = \text{NIAB}$ on it, where N is the number of turns and other symbols take their usual meaning. The spring provides a counter torque $k\theta$ that at equilibrium balances the magnetic torque NIAB resulting in a deflection θ proportional to the current I which is given by;

$$\theta = \left(\frac{NAB}{k}\right)I \tag{2.22}$$

It follows that $\frac{\theta}{I} = \frac{NAB}{k}$

The proportionality constant $\frac{NAB}{k}$ which depends on the design of the meter defines the current sensitivity.

2.1.6 Mass spectrometer

An instrument designed to separate ions according to their charge-to-mass ratio is called mass spectrometer. One type of these instruments is the Bainbridge mass spectrometer shown in Figure 2.15.

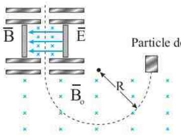


Figure 2.15 Bainbridge mass spectrometer

In this instrument, ions are produced by a source and introduced into a velocity selector consisting of fields $\vec{E}$ and $\vec{B}$. The electric field $\vec{E}$ and magnetic fields $\vec{B}$ are adjusted in such a way that the electric force $F_E = qE$ and magnetic force $F_B = qvB$ acting on the charged particle moving with a specific velocity are balanced, that is $F_E = F_B$ and therefore:

$$v = \frac{E}{B} \qquad (2.23)$$

Consequently, a particle with this velocity is not deflected as it moves through the selector, but other particle moving with different velocity will be deflected. All these ions emerge with the same speed v as they enter a uniform magnetic field B_0 where they travel in a circular path of radius R. In this region the centripetal force is provided by the magnetic force:

$$\frac{mv^2}{R} = qvB_o$$

$$\frac{mv}{R} = qB_o \qquad (2.24)$$

Substituting equation (2.23) into equation (2.24) and rearranging terms we obtain the desired charge to mass ratio for the ions given by:

$$\frac{q}{m} = \frac{E}{RBB_o} \qquad (2.25)$$

Since most ions are singly charged $\left(q = 1.6 \times 10^{-19}\ C\right)$, use of value of R in equation (2.25) makes it possible to determine the mass of ions.

Example 2.10

The strengths of the magnetic and electric fields in the velocity selector of a Bainbridge mass spectrometer are $B = 0.72\ \text{T}$ and $\text{E} = 1.6 \times 10^5\ \text{V/m}$, respectively and the strength of the magnetic field that separates the ions $B_o = 0.88\ \text{T}$ A stream of singly charged lithium ions is found to bend in a circular arc of radius 2.92 cm. Calculate the mass of the lithium ions.

Solution

The mass m of lithium ions can be determined using the charge to mass ratio equation as given by

$$\frac{q}{m} = \frac{E}{RBB_o} \Rightarrow m = \frac{qRBB_o}{E}$$

Thus,

$$m = \frac{1.6\times10^{-19}\,\text{C}\times0.029\,\text{m}\times0.72\,\text{T}\times0.88\,\text{T}}{1.6\times10^{5}\,\text{Vm}^{-1}}$$

$$=1.84\times10^{-26}\ \text{kg}$$

Example 2.11

In the Bainbridge mass spectrometer the magnitude of the magnetic field in the velocity selector is $0.85\,T$ and ions having a speed of 1.62×10^{6} m/s pass through undeflected.

(a) Calculate the magnitude of the electric field in the velocity selector.

(b) If the separation of plates is 6.7 mm, what is the potential difference between plates on the left and right sides?

Solution

(a) For an ion of charge q to be un-deflected, the magnitude of the magnetic force (Bqv) must equal the magnitude of the electric force (qE) and the two forces must have opposite directions, hence

$$qE = Bqv \Rightarrow E = vB$$

$$E = 0.85\,\text{T}\times1.62\times10^{6}\,\text{ms}^{-1}$$

$$=1.38\times10^{6}\ \text{V/m}$$

(b) For a uniform electric field

$$E = \frac{V}{d} \Rightarrow V = Ed$$

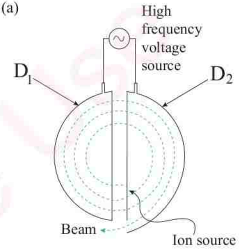
$$V = 6.7\times10^{-3}\,\text{V}\,\text{m}^{-1}\times1.38\times10^{6}\ \text{m}$$

$$V=9.25\times10^{3}\,\text{V}$$

2.1.7 Cyclotron

A cyclotron is a device designed to accelerate charged particles to very high speeds. It was invented by E. O. Lawrence and M. S. Livingston in 1934. A schematic diagram of a typical cyclotron shown in Figure 2.16 (a) consists of an ion source and two semi-circular containers D_1 and D_2, referred to as *dees*, because of their shape resemblance to that of letter D.

(a)



(b)

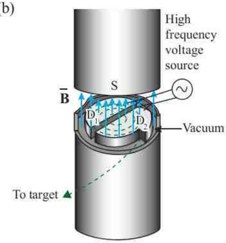


Figure 2.16 (a) A cyclotron and (b) Essential parts of a cyclotron

A high frequency alternating potential difference is applied across the dees to provide electric field.

The dees are enclosed in a steel box and placed between the poles of an electromagnet that provides a uniform magnetic field as shown in Figure 2.16(b). A positive ion is injected with low velocity from an ion source near the center of the device between the *dees*.

It takes half a period $\left(\frac{T}{2}\right)$ to traverse D_1 in a semicircle exclusively due to the magnetic field since the electric field inside the metallic *dees* is zero and arrives at the gap between the *dees*. At $t = \frac{T}{2}$; the polarity of the dees is reversed so that the E-field at the gap accelerates the ion towards D_2. As a result the ion gains kinetic energy qV which in turn increases the orbital radius. In the second half period $\frac{T}{2}$, the ion again arrives at the gap after traversing dee D_2.

The acceleration process is repeated by the varying the polarity of the alternating voltage across D_1 and D_2. When the radius of the orbit becomes equal to the size of the device, high voltage electrodes supply electric field which deflects the ion towards a target.

An expression for the kinetic energy of the ion when it hits the target, in terms of the radius R of the *dees*, can be obtained from equation (2.24) as given by, $v = \frac{RqB}{m}$ and hence, the maximum kinetic energy of the ion incident on the target is

$$KE_{\max} = \frac{1}{2}mv^2 = \frac{(RqB)^2}{2m} \qquad 2.26$$

The operation of the cyclotron as discussed assumes the period T of the ion circulating in the magnetic field is independent of its speed and orbital radius. This is true only for speeds much less than the speed of light.

For higher speeds, relativistic effects takes place such that the period T is no longer independent on speed and that ions do not remain in phase with the applied potential difference needed for acceleration.

One way to overcome this problem is to modify the period of the applied potential difference to keep it in phase with the moving ions. Cyclotron facilities are used to produce radioisotopes needed for different applications in agriculture, anthropology, industry, scientific, research and medicine.

Example 2.12

A cyclotron oscillator's frequency is 10 MHz and the radius of its 'dees' is 50 cm. Determine:

(a) The magnetic field used to accelerate the protons.

(b) The kinetic energy of the accelerated protons.

Solution

From table of constants, charge of proton, $e = 1.6\times10^{-19}\,C$, mass of proton, $m_p = 1.67\times10^{-27}$ kg

(a) From equation (2.24),

$$\frac{mv^2}{r} = qvB \text{, rearranging terms}$$

$$2\pi\, fm = qB \Rightarrow = \frac{2\pi \text{ fm}}{\text{q}}$$

$$B = \frac{2\pi\times10^{7}\text{Hz}\times1.67\times10^{-27}\text{kg}}{1.6\times10^{-19}\text{C}}$$

$$= 0.66 \text{ T}$$

(b) Kinetic energy,

$$KE = \frac{1}{2}mv^2 = \frac{1}{2}m\left(2\pi\, fr\right)^2$$

$$KE = \frac{1}{2}\times1.67\times10^{-27}\,\text{kg}\left(2\pi\times10^{7}\,\text{Hz}\times0.5\,\text{m}\right)^2$$

$$\therefore\ KE = 8.24\times10^{-13}\text{J or } 5.2 \text{ MeV}$$

Exercise 2.1

1. A cosmic ray proton moving toward the earth at 3×10^{7} m/s experiences a magnetic force of 1.3×10^{-18} N. Calculate the strength of the magnetic field if it makes an angle of 65^o with the proton's velocity.
2. An electron moving through a uniform magnetic field $\bar{B} = \left(B_x\hat{i} + 3B_x\hat{j}\right)$T with a velocity $\vec{v} = \left(2\hat{i} + 4\hat{j}\right)$m/s experiences a magnetic force of $F = \left(6.4\times10^{-19}\,N\right)\hat{k}$. Find B_x
3. A cosmic ray electron moves at a speed of 6.5×10^{6} m^{-1} perpendicular to the earth's magnetic field at an altitude where the field strength is 7.0×10^{-6} T. Calculate the radius of the circular path traced by the electron traces.
4. An electron moves at a speed of 3×10^{7} m^{-1} perpendicular to a magnetic field of magnitude 6×10^{-4}T Calculate:
   (a) The radius of the path traced by the electron
   (b) The frequency of revolution of the electron
   (c) The kinetic energy of the electron. (Given: $1\,eV = 1.6\times10^{-19}$ J, mass of electron $= 9.11\times10^{-31}$kg and electron charge $= 1.6\times10^{-19}$C)
5. A wire of length 50 cm and mass 10 g is suspended by a pair of flexible leads in a uniform magnetic field of magnitude 0.5 T as shown in Figure 2.17. Determine the magnitude and direction of the current in the wire required to overcome the tension in the supporting leads.

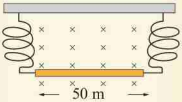


Figure 2.17 Suspended wire

6. A *d.c.* power line for a light-rail system carries 1200 A at an angle of 34º to the earth's magnetic field of magnitude 6.2×10^{-5} T. Calculate the force on a 100 m section of the power line.
7. A proton has a magnetic field due to its spin. The field is similar to that created by a circular current loop 6.5×10^{-6} Å in radius with a current

of 1.05×10^4 A. Find the maximum torque on a proton in a magnetic field of strength 2.5 T.

8. A rectangular current loop of area 24 cm^2 carries a current of 10 A. Calculate the magnetic dipole moment of the loop.
9. A cyclotron used to accelerate alpha-particles of mass 6.64×10^{-27} kg and charge $q = 3.2 \times 10^{-19}$ C has a radius of 0.6 m and magnetic field of strength 2 T. Determine:
   (a) the period of revolution of the alpha-particles; and
   (b) the maximum kinetic energy of the alpha-particles.

2.2 Sources of magnetic fields

The origin of the magnetic field produced by arbitrary distributions of electric current can be explored by the application of the Biot- Savart law. Furthermore, the forces of interaction between two current-carrying wires which leads to the definition of the ampere as a unit of current can be examined. Finally Ampère's circuital law as a simpler alternative to the Biot-Savart law for calculating the magnetic field of symmetrical current configurations will be discussed.

2.2.1 Biot-Savart law

Following the discovery by Oersted that moving charges produce magnetic field and also interact with that field, Jean-Baptiste Biot (1774 – 1862) and Félix Savart (1791–1841) conducted an experiment to investigate the relation between current in a conductor and the magnetic field it produces at a point. The experiment was designed to investigate how the magnetic field $d\vec{B}$ at a point is influenced by varying the current element $Id\vec{l}$, position $\vec{r}$ and orientation θ as illustrated in Figure 2.18.

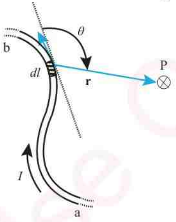


Figure 2.18 Magnetic field $d\vec{B}$ produced by a current element $Id\vec{l}$ at point P

From their investigation, Biot and Savart made several observations: Firstly, the magnetic field $d\vec{B}$ is perpendicular to both the current element $Id\vec{l}$ and the position vector $\vec{r}$. Secondly, the magnitude of the magnetic field dB is proportional to the steady current I, the length dl of the element the sine of the angle θ, and it is inversely proportional to the square of the distance r. Mathematically is given by;

$$dB \propto \frac{Idl\sin\theta}{r^2} \qquad (2.27)$$

This relation is Biot-Savart law. This law could also be expresses as

$$dB = K\frac{Idl\sin\theta}{r^2} \qquad (2.28)$$

The proportionality constant K depends on the medium in which the conductor is situated and is given by $K = \frac{\propto_o}{4\pi}$ where

$\propto_o = 4\pi \times 10^{-7}\, TA^{-1}m$ is the permeability of free space or $K = \dfrac{\propto}{4\pi}$, where $\propto$ is the permeability of a medium.

Relative permeability $\propto_r$ is the ratio of permeability of material to that of free space.

That is, $\propto_r = \dfrac{\propto}{\propto_0}$. It follow that $\propto = \propto_0 \propto_r$

In vector form equation (2.28) is given by

$$d\bar{B} = \frac{\propto_O}{4\pi}\frac{I\left(d\bar{l} \times \bar{r}\right)}{r^3} \qquad (2.29)$$

The magnetic $\vec{B}$ field at point P due to a length of current-carrying wire between *a* and *b* is found by integration.

$$\bar{B} = \frac{\propto_O}{4\pi}\int_a^b \frac{I\left(d\bar{l} \times \bar{r}\right)}{r^3} \qquad (2.30)$$

This integral is rather difficult to evaluate since the integrand is a cross product and therefore contributions from different current elements may not point in the same direction. This limitation can be resolved for a simple configuration by using the following steps: Draw the current element length $Id\bar{l}$ and position vector $\bar{r}$, calculate the cross product $d\bar{l} \times \bar{r}$, use equation (2.30) to solve for the magnetic field and the right-hand rule can be used to verify the direction of the magnetic field.

2.2.2 Magnetic field along the axis of a circular coil

Consider a point P located at a distance x along the axis of a circular coil of radius R carrying a current I as shown in Figure 2.19. The magnetic flux density at the point P can be calculated with the help of Biot-Savat law.

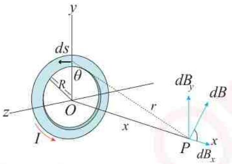


Figure 2.19 Magnetic field along the axis of a circular coil

Using the right hand grip rule, it is noted that the direction of the magnetic field $d\bar{B}$ is perpendicular to the plane formed by $\bar{r}$ and $d\bar{l}$ as shown in Figure 2.19. This vector can be resolved into components dB_x along the x axis and dB_y perpendicular to the x axis. By symmetry, the y-components of the fields, dB_y cancel out as they add up to zero. Each current element $Id\bar{l}$ creates x component of magnitude $dB_x = dBcos\theta$. In this configuration $d\bar{l}$ and $\bar{r}$ are perpendicular so that the magnitude of the vector product $\left(d\bar{l} \times \bar{r}\right)$ in equation (2.29) is rdl and $r^2 = x^2 + R^2$, hence the magnitude of the magnetic field is $dB = \dfrac{\propto_o}{4\pi}\dfrac{I}{\left(x^2 + R^2\right)}dl$.

Summing these components around the entire circular loop and using equation (2.29) for gives:

$$B_x = \oint dB\cos\theta = \frac{\propto_o I}{4\pi}\oint \frac{dl\cos\theta}{x^2 + R^2}$$

where $\cos\theta = \dfrac{R}{r} = \dfrac{R}{\left(x^2 + R^2\right)^{1/2}}$ and hence,

$$B_x = \frac{\propto_o IR}{4\pi\left(x^2+R^2\right)^{3/2}}\oint dl$$

$$= \frac{\propto_o IR^2 N}{2\left(x^2+R^2\right)^{3/2}} \qquad (2.31)$$

where $\oint dl = 2\pi RN$ is the circumference of the coil where N is number of turns of the coil.

Special case 1

If $x=0$ in equation (2.31), point P in Figure 2.19 coincides with the centre of the circular loop. Hence the magnetic field at the centre of the loop $(r=R)$ becomes

$$B = \frac{\propto_o IN}{2R} \qquad (2.32)$$

The magnetic field lines due to a circular wire form closed loops as shown in Figure 2.20. The direction of the magnetic field is given by the right-hand rule, that is, curl the palm of the right hand around the circular wire with the fingers pointing in the direction of the current. The thumb gives the direction of the magnetic field.

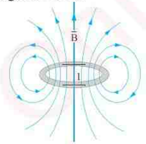


Figure 2.20 The direction of magnetic field for a current loop

Special case 2

For $x >> R$ neglecting the term R^2 in equation (2.31) gives;

$$B = \frac{\propto_o INR^2}{2x^3} \qquad (2.33)$$

The magnitude of the magnetic moment $\propto$ of a loop is defined as the product of current and loop area, $\propto = \mathrm{I}\pi\mathrm{R}^2$ For the circular loop, equation (2.31) can be written in terms of $\propto$ as

$$\mathrm{B} = \frac{\propto_o}{2\pi}\frac{\propto}{x^3} \qquad (2.34)$$

Equation (2.34) is similar to an expression for the electric field due to an electric dipole.

2.2.3 Magnetic field due to a long straight conductor

Consider a long straight conductor carrying a current I as shown in Figure 2.21. The magnetic flux density at point P a distance R from the conductor can be calculated using Biot-Savat law as preceded below.

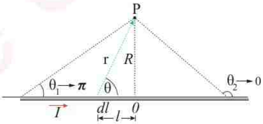


Figure 2.21 Magnetic field due to a long straight conductor

It has been noted that a current loop shown in Figure 2.20 creates a magnetic field similar to that of a bar magnet. A circular conductor is simply a straight conductor bent to form a circular loop. In this section Biot-Savart law is used to determine the magnetic field strength at a point due to a long straight conductor shown in Figure 2.21. Consider the magnetic field at the point P due to the current element $I\overline{dl}$.

Determination of the magnetic field is done by evaluating the vector product in Equation (2.29), $\left|\overline{dl}\times\vec{r}\right| = r\sin\theta dl$.

Thus;

$$B = \frac{\propto_o I}{4\pi}\int \frac{\sin\theta dl}{r^2} \qquad (2.35)$$

Since;

$\sin\theta = \frac{R}{r}$, $r = R\cos ec\theta$ and

$\tan\theta = \frac{R}{l}$, $l = R\cot\theta$ then

$$dl = -R\cos ec^2\theta d\theta$$

$$B = \frac{\propto_o I}{4\pi}\int_{\pi}^{0} \frac{\sin\theta dl}{r^2}$$

$$= \frac{\propto_o I}{4\pi}\int_{\pi}^{o} \frac{\sin\theta\left(-R\cos ec^2\theta\right)d\theta}{\left(R\cos ec\theta\right)^2} \qquad (2.36)$$

The conductor is symmetrical about point O, so setting the limits of integration are set from π to 0, gives:

$$B = \frac{\propto_o I}{2\pi R} \qquad (2.37)$$

2.2.4 Magnetic force between two parallel conductors

Consider two parallel conductors separated by a distance a and carrying currents I_1 and I_2 as shown in Figure 2.22. The magnetic force on either conductor can be determined.

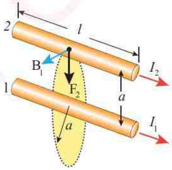


Figure 2.22 Two parallel current-carrying conductors

It was noted that a current-carrying conductor has a magnetic field that obeys the Biot-Savart law. If the current-carrying conductor is placed in external magnetic field it will experience magnetic force whose direction can be determined by the left hand rule. It is evident, that two current-carrying conductors placed in close proximity will exert magnetic forces on each other. André Ampère (1775–1836) investigated the dependence of this force on the magnitude of the current, geometry of the conductors and distance between them. The force between two parallel current-carrying conductors can be used as the basis for defining the unit of current, which is the ampere.

In this section the interest is to determine the force exerted on one conductor as shown in Figure 2.22 due to the magnetic field of the other conductor. The current I_1 in wire 1 sets up a magnetic field $\vec{B}_1$ at the location of wire 2, likewise the current I_2 in wire 2 sets up a magnetic field $\vec{B}_2$ at the location of wire 1. The magnetic fields $\vec{B}_1$ and $\vec{B}_2$ are perpendicular to conductors 2 and 1, respectively.

Using equation (2.16), the magnetic force on conductor 2 is $\vec{F}_2 = I_2\vec{l}\times\vec{B}_1$ and its magnitude $F_2 = I_2lB_1$ since $\vec{B}_1$ is perpendicular to $\vec{l}$. But the magnitude of the magnetic field $B_1 = \frac{\propto_o I_1}{2\pi a}$, Then magnetic force is

$$F_2 = \frac{\propto_o I_1 I_2}{2\pi a}l \qquad (2.38)$$

The magnetic forces between two parallel current-carrying conductors similar to the ones shown in Figure 2.22 are used to define the ampere as a unit of current. It is clear that the magnitudes of the forces per unit length are the same on both conductors, so that equation (2.38) can be written in a more general form as:

$$\frac{F}{l} = \frac{\propto_o I_1 I_2}{2\pi a} \qquad (2.39)$$

For long, parallel conductors separated by 1 metre with each carrying current of 1 A, the force per unit length is:

$$\frac{F}{l} = \frac{(4\pi \times 10^{-7}\,\text{T.m/A})(1\text{A})(1\text{A})}{(2\pi)(1\,\text{m})}$$

$$= 2 \times 10^{-7}\ \text{N/m}$$

From this it is evident that, *ampere* is the current which is flowing in each of two infinitely-long parallel straight conductors of negligible cross-section area separated by a distance of 1metre in vacuum, which produces a force per unit length between the conductors of $2 \times 10^{-7}\,\text{N/m}$.

Activity 2.2

Use the right hand grip rule and Fleming's left hand rule to show that the conductors shown in Figure 2.22 will attract each other when the direction of the current is the same and repel each other if the direction of one of the currents is reversed.

2.2.5 The Ampere's law

André Ampère (1775–1836) formulated a quantitative theorem for calculating the magnetic force exerted by one current-carrying conductor on another. The theorem is an alternative form of Biot-Savart law. It relates the current through a closed path of any shape (an amperian loop) to the magnetic field it produces. The general form of the Ampere's circuital law states that, "*The line integral of $\vec{B} \cdot \overline{dl}$ around any closed path is equal to $\propto_o I$ where I is the total steady current enclosed through any surface bounded by the closed path.*" Mathematically:

$$\oint \text{B} \cdot \overline{dl} = \mu_o \text{I} \qquad (2.40)$$

Equation (2.40) holds well for steady current configurations with high degree of symmetry. The following steps are used when calculating the magnetic field created from current-carrying conductors using Ampere's law: identify the symmetry of the current in the conductor, use the right-hand rule to determine the direction of the magnetic field, select a closed loop where the magnetic field is either constant or zero, calculate the current inside the closed loop, perform the line integral $\oint \bar{B} \cdot \overline{dl}$ around the closed loop and equate with $\propto_o I$ to solve for the magnitude of $\bar{B}$

2.2.6 Applications of Ampere's circuit law

Ampere's law can be used in many ways. For instance can be used to simplify problems with a certain symmetry such as follows:

(a) Magnetic field strength due to a straight wire of radius R carrying a current I.

Consider, an infinitely long straight conductor of radius R carrying a current I_o as shown in Figure 2.23. The magnetic field strength, B at a distance b from the center of the wire for $b \geq R$ can be determined by assuming a uniform distribution of current across the conductor.

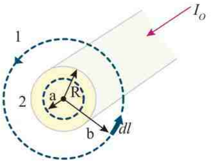


Figure 2.23 Magnetic field due to a straight wire

For loop 1, $b > R$, the enclosed current I_0 is chosen as path of integration as shown in Figure 2.23. By symmetry the magnitude of $\vec{B}$ must be constant and parallel to $d\vec{l}$ at every point on the loop. Therefore, the magnitude of the magnetic field can be determined by integration:

$$\int_0^{2\pi b} \vec{B} \cdot d\vec{l} = Bl\,\Big|_0^{2\pi b} = B(2\pi b) \qquad (2.41)$$

By using the Ampere's circuital law:

$$\oint \mathrm{B} \cdot \overline{dl} = \mu_o \mathrm{I}$$

The expression for the magnetic field in the region $b > R$ is obtained to be;

$$B = \frac{\propto_O I_O}{2\pi b} \qquad (2.42)$$

For the loop 2, $a < R$ the current passing through the plane is less than I_0. Since current is uniformly distributed over the cross-section of the wire, the current enabled by any loop must be proportional to the area it encloses.

That is:

$$I \alpha A$$

$$I = kA \Rightarrow \frac{I}{A} = \frac{I_o}{A_o}$$

$$I = \frac{AI_o}{A_o} = \frac{a^2}{R^2} I_0 \qquad (2.43)$$

Then using the Ampere's law

$$\int_0^{2\pi a} \vec{B} d\vec{l} = B(2\pi a) = \propto_o I$$

$$B(2\pi a) = \frac{\propto_o a^2 I_o}{R^2}$$

When the terms are re-arranged, the magnetic field strength for the region $a < R$ is obtained as:

$$B = \left(\frac{\propto_o I_o}{2\pi R^2}\right) a \qquad (2.44)$$

This result is the same as the one obtained for a long straight conductor using Biot–Savart law in equation (2.37).

The magnetic field outside the conductor $(b \geq R)$ is inversely proportional to the radial distance b (equation (2.42)) while inside the conductor $(a < R)$ is directly proportional to the radial distance a (equation (2.44)). The variation of magnetic field for the above conditions is summarized in Figure 2.24.

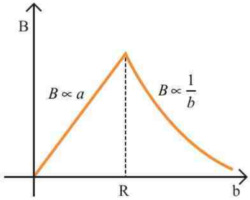


Figure 2.24 Variation of magnetic field

(b) Magnetic field strength, B inside a long solenoid

A solenoid is a long cylindrical coil having many turns. The magnetic field near its centre is uniform and directly proportional to the current in the conductor. This makes solenoid to have wide applications including the cathode ray tube based television, loud speakers, and synchrotron for experimental research. A solenoid has a high symmetrical configuration hence Ampere's circuital law can be conveniently applied to determine the magnetic field strength inside.

Figure 2.25 (a) is a sketch of the magnetic field lines of a loosely wound solenoid. Magnetic field lines close to the wires are similar to those of a long straight current-carrying conductor and are nearly parallel in the interior. If the turns are closely spaced as shown in Figure 2.25 (b), the field lines inside are uniform and parallel to the solenoid axis.

Outside the solenoid, the magnetic field is very weak compared to the field inside and may be assumed to be zero as there are very few lines of force outside, as shown in Figure 2.25(a), the same number of field lines that are concentrated inside spread out into vast space outside the solenoid. Note that the magnetic field lines of a closely wound solenoid resembles that of a bar magnet.

Consider a rectangular path of length l and width w shown in Figure 2.25(b) whose sides 1, 2, 3 and 4 lie on a uniform field. Apply Ampère's law along paths 1 to 4 of the rectangular loop by evaluating $\int \vec{B}\cdot dl$ over each side of the rectangle. Since $\vec{B}$ is uniform and parallel to $\vec{dl}$ the contribution of path 1 is from S to R,

$$\int_S^R \vec{B}\cdot\vec{dl} = Bl$$

The magnetic field $\vec{B}$ and elemental length $\vec{dl}$ for both paths 2 (from R to Q) and 4 (from P to S) are perpendicular and hence their dot product are zero. Then:

$$\int_R^Q \vec{B}\cdot\vec{dl} = \int_P^S \vec{B}\cdot\vec{dl} = 0$$

The integral $\int_Q^P \vec{B}\cdot\vec{dl}$ along path 3 (from Q to P) is zero because the magnetic field outside the solenoid is negligible. The integral over the closed rectangular path is therefore:

$$\oint \vec{B}\cdot\vec{dl} = \int_S^R \vec{B}\cdot\vec{dl} + \int_R^Q \vec{B}\cdot\vec{dl} + \int_Q^P \vec{B}\cdot\vec{dl} + \int_P^S \vec{B}\cdot\vec{dl} = Bl \qquad (2.45)$$

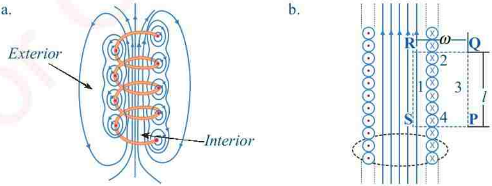


Figure 2.25 (a) The magnetic field lines for a loosely wound solenoid and (b) cross-section of a tightly wound solenoid.

Equating this result with $\propto_o \mathrm{NI}$, where N is the number of turns in the solenoid, gives:

$$Bl = \propto_o \mathrm{NI}$$

$$B = \frac{\propto_o \mathrm{NI}}{l} = \propto_o \mathrm{nI} \qquad (2.46)$$

where n is the number of turns per unit length.

From equation (2.46) it is noted that the magnetic field strength B depends on the permeability of material of the core which in this case is permeability of free space $\propto_o$ and directly proportional to the product of number of turns per unit length n and the current I.

(c) Magnetic field strength, B inside and outside a toroid

A toroid is a donut-shaped closely wound coil of wire as shown in Figure 2.26.

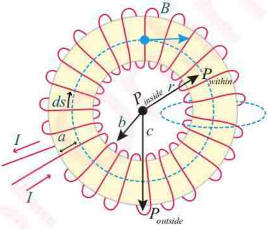


Figure 2.26 A toroid

A toroid can be considered as a solenoid twisted around to have its ends meet. Thus, the magnetic field is completely confined inside the coil that forms the toroid. Applying Ampere's circuital law to Figure 2.26.

(i) *$\vec{B}$ at a point outside the toroid*

Consider an Amperian loop passing through $P_{outside}$;

$$\oint B\,dl = \propto_o I$$

But the net current in this loop is zero

$$\oint B\,dl = 0$$

$$BL = 0$$

$$B = 0$$

(ii) *$\vec{B}$ at a point inside the toroid*

Consider the Amperian loop passing through P_{Inside}

$\oint B\,dl = \propto_o I$ but I is again zero

$\therefore B = 0.$

(iii) *$\vec{B}$ at a point within the toroid*

If an Amperian loop passing through point P_{within} is considered;

$$\oint B \cdot dl = \oint B\,dl\cos\theta = \propto_0 I$$

But; $\theta = 0\Upsilon$

$$\oint B\,dl = \propto_o I \Rightarrow BL = \propto_o I$$

$$\therefore B = \frac{\propto_o I}{L} = \frac{\propto_o I}{2\pi r}$$

It is noted that the magnetic field inside the toroid is non-uniform and varies as $\frac{1}{r}$. Since a toroid is an endless solenoid the length, l of the solenoid is replaced by $2\pi r$.

Exercise 2.2

1. A circular coil with 200 turns has a radius of 5 cm and carries a current of 5 A. Calculate the magnetic field strength at:
   (a) the centre of the coil
   (b) a point at a distance of 10 cm along the axis of the coil.
2. A toroid has a core of inner radius 18 cm and outer radius 20 cm around which 300 turns of the wire are wound. What is the magnetic field of the toroid if a current of 5 A flows in the wire?
3. What is the magnetic field strength at the centre of a semicircular piece of wire with radius 0.10 m and carrying 20 A of current?
4. What is the magnetic field strength at point P due to the current I flowing in the wire as shown in Figure 2.27

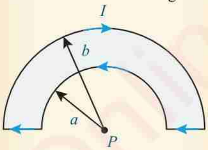


Figure 2.27 A wire

5. Two long, straight parallel wires are 25 cm apart.
   (a) If each wire carries a current of 50 A in the same direction, what is the magnetic force per metre exerted on each wire?
   (b) Are the forces repulsive or attractive?
   (c) What would happen if the currents flow in opposite directions?

2.3 Electromagnetic induction

Nowadays, currents induced by magnetic fields are essential to our technological society. Many devices including electric generators, transformers, pickup coils in electric guitars, some types of microphones, tablet computers, Automated Teller Machines (ATM) cards, electric vehicles and public announcement loud speakers use magnetic induction for their operations. In the discussion of magnetic fields produced by constant currents it is assumed that the fields are time independent. Most of the important applications of electromagnetism utilize time dependent magnetic fields. The phenomenon in which electric current is generated by changing the magnetic flux is known as *electromagnetic induction*.

2.3.1 Magnetic flux

Magnetic flux, Φ_B is a measure of the strength of the magnetic field at a point. It is a useful tool for describing the effects of the magnetic force on a conductor occupying a given area. Magnetic flux through a plane of area A with normal vector $\hat{n}$ placed in a uniform magnetic field B as shown in Figure 2.28 can be written as:

$$\Phi_B = \vec{B} \cdot \vec{A}\hat{n} = BA\cos\theta \qquad (2.48)$$

where θ is the angle made between the normal to the surface area and the direction of the magnetic field $\vec{B}$.

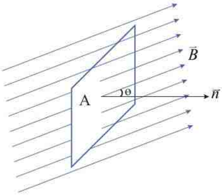


Figure 2.28 Unit vector n placed in a uniform magnetic field B

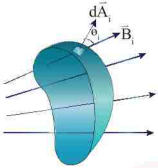


Figure 2.29 Magnetic field $\vec{B}_i$ at the i^{th} area element, $d\vec{A}_i$

Special cases for $\Phi_B = BA\cos\theta$

(i) If the uniform field $\vec{B}$ makes an angle $\theta = 0^o$ with $\hat{n}$, then $\Phi_B = BA\cos 0^o = BA$

(ii) If the uniform field $\vec{B}$ makes an angle $\theta = 90^o$ with $\hat{n}$, then $\Phi_B = BA\cos 90^o = 0$

If the magnetic field has different magnitudes and directions at different regions on a surface as shown in Figure 2. 29, then the magnetic flux through the surface is given by a summation through the whole area, given by:

$$\Phi_B = \sum_{i=1}^{n} \vec{B} \cdot \Delta\vec{A}_i, \text{where} \quad i = 1,2,3,4....,n \tag{2.49}$$

In the limit as $\Delta\vec{A}_i \rightarrow 0$, the summation becomes an integral over the entire region and gives the exact value of a magnetic flux as;

$$\Phi_B = \int_A \vec{B} \cdot d\vec{A} \tag{2.50}$$

The SI unit of magnetic flux is Weber (Wb). If $B = 1\,T$ and $A = 1\,m^2$, then $\Phi = 1T \times 1\,m^2 = 1\,Wb$, hence 1 Weber is the amount of magnetic flux passing through an area of $1\,m^2$ normal to the uniform magnetic field of 1 T.

2.3.2 Laws of electromagnetic induction

Michael Faraday demonstrated that, when the magnetic flux through a coil changes with time, an *e.m.f (ε)* is induced. If a coil has multiple turns N, then the *flux linkage* Φ_B through the coil is the sum of individual fluxes in each turn, i.e

$$\Phi_B = NBA\cos\theta \tag{2.51}$$

Faraday's law of electromagnetic induction state that,

"Whenever the magnetic flux in a circuit changes, an e.m.f. is induced whose magnitude is directly proportional to the rate of the change of magnetic flux through a closed circuit".

The above statement is known as Faraday's law of electromagnetic induction it can be expressed mathematically as :

$$\varepsilon = -\frac{d\Phi_B}{dt} \qquad (2.52)$$

Lenz's law of electromagnetic induction

The negative sign in equation (2.52) expresses the Lenz's law which states that *induced current is always such as to oppose the cause producing it.*

For a coil of N turns, using equation (2.52), the induced e.m.f. becomes

$$\varepsilon = -\frac{d\Phi_B}{dt} = -N\frac{d\left(BA\cos\theta\right)}{dt} \qquad (2.53)$$

From this expression, the e.m.f. (ε) can be induced in the circuit by changing with respect to time:

(i) The magnetic field strength $\vec{B}$

(ii) The area enclosed by the closed loop.

(iii)The angle θ between $\vec{B}$ and the unit vector $\hat{n}$ normal to the area of the loop.

Of all the approaches (i) – (iii) that can induce e.m.f., change of the angle by rotation of the loop in a constant magnetic field is the easiest and hence most practical.

2.3.3 Induced e.m.f. in a moving straight conductor in a uniform magnetic field

Consider a straight conductor of length l moving in a uniform magnetic field of strength B with velocity v as indicated in Figure 2.30.

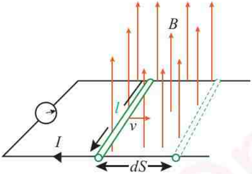


Figure 2.30 Straight conductor moving in a uniform magnetic field

Area swept by the conductor, in time dt is dA Magnetic flux change $d\Phi_B = BdA$. But $dA = ldS$ where dS is the distance moved by the conductor.

$d\Phi = BldS$

With $\varepsilon = \frac{-d\Phi}{dt} = Bl\frac{dS}{dt}$

Thus induced e.m.f., is given as $\varepsilon = Blv$,

where $\frac{dS}{dt} = v$

2.3.4 Induced e.m.f. in a rotating coil

Consider a coil rotating in a uniform magnetic field at an angular velocity ω as shown in Figure 2.31.

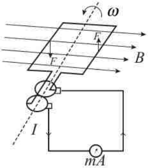


Figure 2.31 Rotating coil in a uniform magnetic field

Magnetic flux linking the coil,

$$\Phi_B = BNA\cos\theta$$

From Faradays and Lenz's law of electromagnetic induction

$$\varepsilon = -\frac{d\Phi_B}{dt}$$

$$\varepsilon = -\frac{dNAB\cos\omega t}{dt}, \text{where } \theta = \omega t$$

$$\varepsilon = -NAB\frac{d\cos\omega t}{dt}$$

$$\varepsilon = \omega NAB\sin\omega t$$

For maximum induced *e.m.f.*,

$$\sin\omega t = 1$$

$$\varepsilon = \omega NAB$$

Example 2.13

The current through a solenoid of length 60 cm and radius 1.5 cm with 1500 turns/metre is changing at a rate of 2 As^{-1}. A small coil consisting of 30 closely wound turns wrapped in a circle of radius 0.5 cm is placed inside of the solenoid in such a way that the plane of the coil is perpendicular to the central axis of the solenoid. Calculate the magnitude of the induced (e.m.f.).

Solution

By definition $\Phi_B = BA$. However for a solenoid $B = \propto_0 In$ and $A = \pi r^2$

r = radius of solenoid $\Phi_B = \propto_0 In\pi r^2$

$\frac{d\Phi_B}{dt} = \propto_0 n\pi r^2 \frac{dI}{dt}$ implies that e.m.f.,

$$\varepsilon = -\propto_0 n\pi r^2 \frac{dI}{dt}$$

For N turns the magnitude of the induced e.m.f. is given as:

$$\varepsilon = \left|N\frac{d\Phi_B}{dt}\right| = \left|N\propto_o n\pi r^2\frac{dI}{dt}\right|$$

$$= 30\times 4\pi\times 10^{-7}\,\text{TmA}^{-1}\times 1500\text{ turns m}^{-1}$$

$$\times\pi\times\left(0.5\times 10^{-2}\,\text{m}\right)^2\times 2\text{ As}^{-1}$$

$$= 8.9\times 10^{-6}\,\text{V}$$

Example 2.14

A uniform magnetic field is normal to the plane of a circular loop 15 cm in diameter and made of copper wire of radius 0.18 cm. Determine the rate of change of the magnetic field if an induced current of 7.3 A is to appear in the loop. Resistivity of copper is $1.68\times 10^{-8}\,\Omega\,\text{m}$.

Solution

The induce *e.m.f.* in the coil is given as

$\varepsilon = IR$ and $R = \frac{\rho l}{A} = \frac{\rho\pi D}{\pi r^2}$ where D and r are the diameter of the coil and radius of the copper wire respectively, hence

$$\varepsilon = \frac{I\rho D}{r^2}$$

$$\varepsilon = \frac{7.3\,A\times 1.68\times 10^{-8}\,\Omega\,\text{m}\times 15\times 10^{-2}\,\text{m}}{\left(0.18\times 10^{-2}\,\text{m}\right)^2}$$

$$= 5.7\times 10^{-3}\,\text{V}$$

Also the induce *e.m.f.*, the rate of change of magnetic field and the area of the loop A_{loop} are related by the equation given as

$$\varepsilon = -A_{loop}\frac{dB}{dt} \Rightarrow \frac{dB}{dt} = \frac{-\varepsilon}{A_{loop}}$$

$$\frac{dB}{dt} = \frac{-\left(-5.7\times10^{-3}\,\text{V}\right)}{\pi\left(\frac{0.15\,\text{m}}{2}\right)^2} = 0.32\ \text{Ts}^{-1}$$

Example 2.15

Show that, a coil of an area 100 cm^2 with 8000 turns and making 600 revolutions per minute in a magnetic field of flux density $5\times10^{-2}\,T$, produces a sinusoidal e.m.f. with amplitude of 251 V.

Solution

In general a sinusoidal *e.m.f.* (ε) is related to the amplitude ε_o and given by:

$$\varepsilon = \varepsilon_o \sin\omega t$$

The angular speed

$$\omega = \left(2\pi\frac{radian}{rev}\right) f\left(\frac{rev}{s}\right)$$

$$\omega = \left(2\pi\right)\left(600\frac{rev}{\text{min}}\right)\left(\frac{1\,\text{min}}{60\,s}\right)$$

$$\omega = 62.8\ \text{rad/s}$$

But $\varepsilon_o = BAN\omega$

$$\varepsilon_o = \left(5\times10^{-2}\text{T}\right)\left(1\times10^{-2}\ \text{m}^2\right)\left(8000\right)\left(62.8\ \text{s}^{-1}\right) = 251\ \text{Tm}^2\text{s}^{-1}$$

Since $1\ \text{V} = 1\ \text{Wbs}^{-1}$ and $1\ \text{T} = 1\ \text{Wbm}^{-2}$, then

$$\varepsilon_o = 251\left(\text{Wbm}^{-2}\right)\text{m}^2\text{s}^{-1}$$

$$= 251\ \text{Wbs}^{-1}$$

$$= 251\ \text{V} \qquad \text{Hence shown}$$

2.3.5 Mutual-induction

Consider two coils, 1 and 2 placed close to each other as shown in Figure 2.32. Suppose current I_1 is allowed to flows in coil 1, a magnetic flux will be set up and link with coil 2. If this current is varied, the flux linked to coil 2 also changes and hence an e.m.f. is induced in coil 2. The e.m.f. induced in coil 2 is mutual induced e.m.f. and the process is termed as mutual induction.

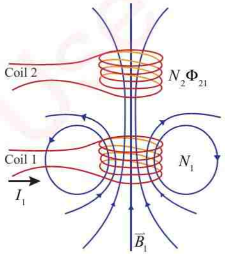


Figure 2.32 Changing current in coil 1 produces a changing magnetic flux in coil 2

Let Φ_{21} denote the magnetic flux through one turn of coil 2 due to current I_1 in coil 1. By varying I_1 the induced e.m.f. associated with changing magnetic flux in coil 2 is given by:

$$\varepsilon_{21} = -N_2\frac{d\Phi_{21}}{dt} = -\frac{d}{dt}\left(\int_{coil\,2}\overline{B}\cdot d\overline{A_2}\right) \quad (2.54)$$

The rate of change of magnetic flux Φ_{21} in coil 2 is proportional to the rate of change of the current I_1 in coil 1. Thus,

$$N_2 \frac{d\Phi_{21}}{dt} = M_{21} \frac{dI_1}{dt} \qquad (2.55)$$

The proportionality constant M_{21} is called the *mutual inductance*. The SI unit of mutual inductance is Henry (H). Integration of equation (2.55) gives:

$$M_{21} = \frac{N_2 \Phi_{21}}{I_1} \qquad (2.56)$$

If the total flux $N_2\Phi_{21} = 1\,Wb$ and the current $I_1 = 1\,A$ then $M_{21} = 1\,H$. Hence, *mutual inductance between two coils is 1 henry (H) if a current of 1 A flowing in one of the coils sets up a total flux of 1 Wb in the other coil.* One henry can also be expressed in terms of the magnetic flux density B in tesla, area of the coil and the current as:

$$1\,H = 1\frac{Tm^2}{A} \qquad (2.57)$$

Considering time varying current I_2 flowing in coil 2 as indicated in Figure 2.33, the induced e.m.f. in coil 1 becomes:

$$\varepsilon_{12} = -N_1 \frac{d\Phi_{12}}{dt} = -\frac{d}{dt}\left(\int_{coil1} \vec{B}\cdot d\vec{A_1}\right) \qquad (2.58)$$

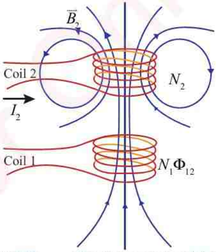


Figure 2.33 Changing current in coil 2 causes the change in magnetic flux in coil 1

The rate of change of magnetic flux ϕ_{12} in coil 1 is proportional to the rate of change of the current I_2 in coil 2

$$N_1 \frac{d\Phi_{12}}{dt} = M_{12} \frac{dI_2}{dt} \qquad (2.59)$$

M_{12} can also be expressed in terms of total flux as

$$M_{12} = \frac{N_1 \Phi_{12}}{I_2} \qquad (2.60)$$

Using equations (2.46), (2.50), (2.56) and (2.60) the mutual inductance M_{21} and M_{12} can be written as:

$$M_{21} = \frac{N_2 \Phi_{21}}{I_1}, \text{ then}$$

$$M_{21} = \frac{N_2 B_1 A_2}{I_1} = \frac{N_2\left(\frac{\propto_o N_1 I_1}{l}\right)A_2}{I_1}$$

$$= N_2\left(\frac{\propto_o N_1}{l}\right)A_2$$

$$M_{21} = \frac{\propto_o N_1 N_2 A_2}{l} \qquad (2.61)$$

Similarly,

$$M_{12} = \frac{N_1 B_2 A_1}{I_2} = \frac{N_1\left(\frac{\propto_o N_2 I_2}{l}\right)A_1}{I_2}$$

$$= N_1\left(\frac{\propto_o N_2}{l}\right)A_1$$

$$M_{12} = \frac{\propto_o N_1 N_2 A_1}{l} \qquad (2.62)$$

From equations (2.61) and (2.62), for coils of the same area and length, the mutual inductance $M_{21} = M_{12}$.

Thus the mutual inductance between the two coils is the same regardless of which coil carries the current and therefore can be given by

$$M = \frac{\propto_0 N_1 N_2 A}{l} \qquad (2.63)$$

From this equation it can be concluded that the mutual inductance depends only on the number of turns and radii of the two coils. Actually the mutual inductance is intimately related to self-induction which is described in the next sub-section.

Example 2.16

Consider two co-axial coils, of radii r_1 and r_2 such that $r_1 << r_2$, carrying current I_1 and I_2 with corresponding number of turns N_1 and N_2 respectively Figure 2.34. Determine the mutual inductance of the two coils.

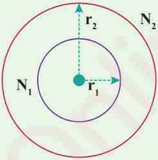


Figure 2.34 *Two co-axial concentric coils*

Solution

The magnetic field at the center due to a current I_2 is given by;

$$B_2 = \frac{\propto_0 N_2 I_2}{2r_2}$$

Since the primary (inner) co-axial coil has a very small radius, B_2 may be considered constant over the cross-sectional area of the primary coil.

Hence, the total magnetic flux linkage with the primary (inner) coil is given by

$$N_1\Phi_1 = N_1 B_2 A_1 = N_1\left(\frac{\propto_0 N_2 I_2}{2r_2}\right)\pi r_1^2$$

$$N_1\Phi_1 = \left(\frac{\propto_0 \pi N_1 N_2 r_1^2}{2r_2}\right) I_2$$

But $N_1\Phi_1 = N_2 I_2$, therefore, the mutual-inductance of the primary coil with respect to the secondary coil is given by;

$$M_{12} = \frac{N_1\Phi_1}{I_2} = \frac{\propto_0 \pi N_1 N_2 r_1^2}{2r_2}$$

It is not easy to calculate the flux linkage with the secondary (outer) coil as the magnetic field due to the primary (inner) coil varies across the cross section of the secondary coil. Therefore, the calculation of M_{21} will also be extremely difficult in this case. The equality $M_{12} = M_{21} = M$ is given by reciprocity theorem which is very useful in such situations. Therefore, the mutual-inductance of the secondary with respect to the primary is given by

$$M_{12} = M_{21} = \frac{\propto_0 \pi N_1 N_2 r_1^2}{2r_2}$$

In its simplest form, the reciprocity theorem used above states that "if an induced *e.m.f.* (ε) in one branch of a reciprocal network produces a current I in another, then if *e.m.f.* (ε) is moved from the first to the second branch, it will cause the same current in the first branch, where the induced *e.m.f.* has been replaced by a short circuit."

Example 2.17

A solenoid of length 70 cm and radius 4.5 cm has 2100 turns. A second coil of 750 turns is wound on the middle of the first solenoid. Use this information to estimate the mutual inductance M between the coils.

Solution

The mutual inductance,

$$M = \frac{\mu_0 N_1 N_2 A}{l}$$

$$M = \frac{\left(4\pi\times10^{-7}\frac{H}{m}\right)(750)(2100)\pi\times(0.045\,\text{m})^2}{(0.7\text{ m})}$$

$$= 1.80\times10^{-2}\text{ H}$$

Example 2.18

The current in the primary coil of a co-axial solenoid is reduced from 3 A to zero in 1 ms. The process induces an *e.m.f.* of 1500 V in the secondary coil. Use the information to estimate the mutual inductance between the primary and secondary coils.

Solution

The magnitude of induced *e.m.f.* (ε) and the mutual inductance M are related by,

$$|\varepsilon| = M\frac{dI}{dt}$$

Where $|\varepsilon| = 1500\text{ V}$, $dI = (3-0)\text{ A} = 3\text{ A}$ and $dt = 0.001\,s$

$$M = \frac{|\varepsilon|}{\frac{dI}{dt}} = \frac{1500\text{ V}}{3\text{ A}}\times0.001\text{ s}$$

$$M = 0.50\text{ VA}^{-1}\text{s} = 0.50\text{ H}$$

2.3.6 Self - inductance

In self-inductance only one coil consisting of N turns and carrying a time varying current I is considered as shown in Figure 2.35.

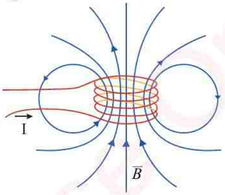


Figure 2.35 Magnetic flux through the current loop

According to Lenz's law, an e.m.f. will be induced in such a way that it opposes the change in the flow of the current. The e.m.f. generated is called *self-induced* e.m.f., and the process is known as *self-induction.*

Mathematically, the self-induced e.m.f. is given as

$$\varepsilon_L = -N\frac{d\Phi_B}{dt} = -N\frac{d}{dt}\left(\int\vec{B}\cdot d\vec{A}\right) \quad (2.64)$$

Since $N\frac{d\Phi_B}{dt} \propto \frac{dI}{dt}$, the e.m.f. becomes

$\varepsilon_L \propto \frac{dI}{dt}$ and hence,

$$\varepsilon_L = -L\frac{dI}{dt} \quad (2.65)$$

The proportionality constant L is called the *coefficient of self-induction* or *self-inductance*. The SI unit of self-inductance is Henry, H.

Combining equations (2.64) and (2.65) and integrating the result and then rearranging the terms gives:

$$L = \frac{N\phi_B}{I} \qquad (2.66)$$

From this equation, self-inductance is essentially flux linkage of the coil per unit current.

Example 2.19

The self-induced *e.m.f.* in an inductor is 17V and the rate of change of the current is 25 kA/s. Calculate the self-inductance.

Solution

The self-induced *e.m.f.* in an inductor is given as

$$\varepsilon = -L\frac{dI}{dt} \Rightarrow L = -\frac{\varepsilon}{(dI/dt)}$$

$$L = -\frac{17\text{ V}}{(-25\times10^3\text{ A/s})} = 0.68\text{ mH}$$

Example 2.20

The current in an inductor of self-inductance 3.4 H is given by $I = 6t^2 + 22t + 47$. Find an expression for the magnitude of the induced (*e.m.f.*).

Solution

Using the equation $\varepsilon = -L\frac{dI}{dt}$ where

$$\frac{dI}{dt} = \frac{d}{dt}(6t^2 + 22t + 47) = 12t + 22$$

then $|\varepsilon| = \left|-L\frac{dI}{dt}\right| = 3.4(12t + 22)$

and hence $|\varepsilon| = (40.8t + 74.8)\text{V}$

Activity 2.3

Design a simple circuit using battery, two bulbs, switch, an inductor and connecting wires, to verify self-induced (*e.m.f.*).

2.3.7 Energy stored in an inductor

An inductor generates induced e.m.f. in a circuit which opposes any change in the current through it. In order to establish a current in an inductor, work must be done by an external source such as a battery against back (e.m.f.). Using the work-energy theorem, the rate at which an external *e.m.f.* ε_{ext} works against the self-induced *e.m.f.* ε_L to allow the passage of current I in the inductor is given as;

$$\frac{dW}{dt} = I\varepsilon_{ext} \qquad (2.67)$$

In absence of other external forces, the external *e.m.f.* ε_{ext} does work only against the back *e.m.f.* ε_L hence $\varepsilon_{ext} = -\varepsilon_L$. This implies;

$$\frac{dW}{dt} = -I\varepsilon_L = -I\left(-L\frac{dI}{dt}\right) = IL\frac{dI}{dt} \qquad (2.68)$$

Upon integration, the total work done by the external field to increase the current from zero to I is

$$W = \int_0^I LI\,dI = -\frac{1}{2}LI^2$$

The work done is essentially a stored magnetic energy in the coil,

$$U_B = \frac{1}{2}LI^2 \qquad (2.69)$$

Example 2.21

A coil of N turns with magnetic flux linkage (Φ), in each turn due to the current (I) passing through it has self-inductance given by $L = \frac{N\Phi}{I}$. Use this definition to show that:

(a) The inductance of the solenoid is given as:

$L = \frac{\propto_0 \propto_r N^2 A}{l}$, where $\propto_r$ is the relative permeability of the core.

(b) The energy stored in the inductor is $U = \frac{1}{2} L I_0^2$ where I_o is the final value of current.

Solution

(a) Inductance of the coil $L = \frac{N\Phi}{I}$ where $\Phi = BA$.

But $B = \frac{\propto NI}{l} \Rightarrow \Phi = \left(\frac{\propto NI}{l}\right) A$

For air-cored coil

$$L = \frac{N\Phi}{I} = \frac{N}{I}\left(\frac{\propto NI}{l}\right) A = \frac{\propto N^2 A}{l}$$

But $\propto_r = \frac{\propto}{\propto_0}$ hence

$$L = \frac{\propto_0 \propto_r N^2 A}{l}$$

(b) If the current rate is $\frac{dI}{dt}$ the magnitude of the induced *e.m.f.* in the inductor

$$\varepsilon = L\frac{dI}{dt}$$

If the current I passing through the inductor in a small time interval dt, then it must be doing work dW given as:

$$dW = \varepsilon I dt = LI\frac{dI}{dt}dt$$

The total work done to increase the current from zero to I_o is given by:

$$W = \int_0^{I_o} LI\,dI = \left.\frac{LI^2}{2}\right|_0^{I_o} = \frac{1}{2} L I_o^2$$

By conservation of energy the work done on the system becomes energy stored in the inductor:

$$U = \frac{1}{2} L I_0^2$$

Example 2.22

A 12 V battery is connected in series with a $30\,\Omega$ resistor and a coil of inductance 220 mH. At some time, the current in the circuit increases to half of its peak value. Use this information to estimate the rate at which:

(a) Energy is being delivered by the battery.

(b) Energy is being stored in magnetic field of the inductor.

Solution

(a) The maximum current possible is:

$$I_o = \frac{\varepsilon}{R} = \frac{12\text{ V}}{30\,\Omega} = 0.4\text{ A}$$

The current that was given at the specified time was $\frac{I_o}{2} = 0.2$ A.

Thus the power supplied by the battery at that time was then;

$$P = \varepsilon I = (12\text{V})(0.2\text{ A}) = 2.4\text{ W}$$

(b) The energy stored is given by:

$U = \frac{1}{2} L I^2$ where I is the current at specified time t.

The rate at which energy is stored in the inductor is

$$\frac{dU}{dt} = \frac{d}{dt}\left(\frac{1}{2} L I^2\right) = LI\frac{dI}{dt}$$

But $L\dfrac{dI}{dt}=\varepsilon=12-IR$ and hence:

$$\frac{dI}{dt}=\frac{\varepsilon}{L}=\frac{12-IR}{L}$$

$$=\frac{12V-(0.2A)(30\Omega)}{0.22\text{H}}$$

$$\frac{dI}{dt}=\frac{6}{0.22}\frac{\text{A}}{\text{s}}$$

$$\therefore \frac{dU}{dt}=(0.22\,\text{H})\left(\frac{6}{0.22}\frac{\text{A}}{\text{s}}\right)(0.2\,\text{A})$$

$$=\left(0.22\frac{\text{kgm}^2}{\text{s}^2\text{A}^2}\right)\left(\frac{6}{0.22}\frac{\text{A}}{\text{s}}\right)(0.2\,\text{A})$$

$$=1.20\frac{\text{kgm}^2}{\text{s}^2}\left(\frac{1}{\text{s}}\right)$$

$$=1.20\frac{\text{J}}{\text{s}}$$

$$=1.20\,\text{W}$$

Exercise 2.3

1. What factors influence the magnitude of induced e.m.f. in an electric circuit?
2. Does a change in magnetic flux result to both induced *e.m.f.* and induced current or induced *e.m.f.* only? Explain.
3. The current in an inductor is changing at the rate of 200 A/s, and the inductor *e.m.f.* is 80 V. What is its self-inductance?
4. A cardboard tube measures 15 cm long by 2.2 cm in diameter. How many turns of wire must be wound on the full length of the tube to make a 5.8 mH inductor?
5. Show that $\dfrac{N\Phi}{I}$ and $\dfrac{\varepsilon dt}{dI}$ which are both expressions for self-inductance, have the same units of *H*.
6. The current (I) flowing in a coil of self-inductance $L = 8.0\ H$ changes with time according to $I = 4.0\sin 2t$. Estimate the energy stored in the inductor as the current increases from $t=0$ to a final time $t=\dfrac{\pi}{4}$ in seconds.
7. (a) Show that the units of e.m.f. in volts are $\dfrac{\text{kgm}^2}{\text{s}^3\,\text{A}}$.

   (b) Calculate the number of turns, N of a solenoid of length 0.12 m and radius 0.02 m given the rate of change of current in the solenoid is $0.8\,\text{As}^{-1}$ and the induced *e.m.f.* is 6.06×10^{-4} V.
8. A 32.0 V *e.m.f.* is induced in a 0.32 H coil by a current I that increases uniformly from zero to I_0 in 0.002 s. Estimate the value of I_0.
9. (a) A toroid of 1200 turns with an air core has a radius of 15 cm and cross sectional area of 12.0 cm^2. What is the self-inductance of the toroid?

   (b) Another coil of 300 turns is wound closely to the toroid in (a) above. If the current in the toroid is changed from zero to 2.0 A in 0.05 s, what e.m.f. is induced in the second coil?
10. Primary and secondary coils of 150 and 200 turns, respectively are wound side by side on a closed iron circuit of cross sectional area of 150 cm^2 and length 300 cm. Calculate:

(a) The mutual inductance between the coils.
(b) The induced *e.m.f.* in the second coil if the current changes from zero to 10 A in the first coil in 0.02 s. Use the relative permeability of iron $\propto_r = 2000$.

2.4 Magnetic properties of materials

Study of properties of magnetic materials will enable the understanding of mechanism behind the magnetic behaviour of materials. Such knowledge may allow us to alter and tailor the magnetic properties of materials to suit specific applications. The knowledge gained will enable you to:

(i) explain the origins of magnetization of different materials
(ii) classfying the magnetic materials into diamagnetic, paramagnetic and ferromagnetic materials.

2.4.1 Origin of magnetism from atomic point of view

Magnetism of solid materials originates from the motion of orbital electrons around the nucleus of atoms. Consider an electron of charge e and mass m moving in a circular orbit with radius r and speed v as shown in Figure 2.36.

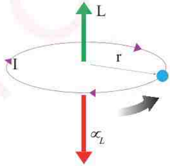


Figure 2.36 Orbital motion of an electron around the nucleus of an atom

This moving charge is equivalent to a current loop which, as described in previous sections, has a magnetic dipole moment $\propto_L = IA$ where $A = \pi r^2$ is the area enclosed by the loop carrying current I. The equivalent current I is the total charge passing at any point on the orbit per unit time, which is just the magnitude of the electron charge divided by the orbital period $T = \frac{2\pi r}{v}$ and is expressed as:

$I = \frac{e}{T} = \frac{ev}{2\pi r}$, where v is the linear velocity of electron.

The magnetic moment $\propto_L$ as a product of current and area becomes:

$$\propto_L = \frac{ev}{2\pi r}\left(\pi r^2\right) = \frac{evr}{2} \qquad (2.70)$$

It is useful to express μ in terms of *angular momentum* L of the electron. For a particle moving in a circular path, the angular momentum is $L=mvr$; equation 2.70 can then be written in terms of angular momentum as follows:

$$\propto_L = \frac{e}{2m}L \qquad (2.71)$$

This equation demonstrates that the magnetic moment of the electron, which is the origin of magnetism in matter, is proportional to its orbital angular momentum. However all substances contain electrons, and yet only a few are magnetic. The main reason is that in most substances, the magnetic moment of one electron in an atom is canceled by that of another electron orbiting in the opposite direction. Consequently, the magnetic effect produced by the orbital motion of the electrons for these materials is either zero or very small.

2.4.2 Magnetization

In the previous section it is noted that orbiting electrons in an atom have magnetic moments. In matter these moments may add to give zero or non-zero net magnetic moments. The magnetic state of a substance is described by a quantity called the magnetization M defined as the net magnetic moment per unit volume. It is a vector quantity given by:

$$\vec{M} = \frac{\vec{\propto}_L}{V} \qquad (2.72)$$

When an external magnetic field $\vec{B}_0$ is applied to a material, the material gets magnetized. The total magnetic field inside the material is the sum of the applied magnetic field and that created by magnetization of the material, expressed as.

$$\vec{B} = \vec{B}_0 + \vec{B}_m \qquad (2.73)$$

where $\vec{B}_m$ is the magnetic field due to magnetization of the material produced by spinning of electrons. Since $\vec{B}_0 = \propto_0 \vec{H}$ where, $\vec{H}$ is called magnetic intensity or magnetic field intensity

$$B_m = \propto_0 H_m$$
$$B = \propto_0 H + \propto_0 H_m$$
$$B = \propto_0 \left(H + H_m\right)$$

From equation (2.72) $\vec{M}$ in vector notation can be expressed as:

$$\vec{M} = \frac{NAI_m}{Al} = \frac{N}{l} I_m$$

$\overline{M} = nI_m$ where $n = \dfrac{N}{l}$, number of turns per unit length

But $H_m = nI_m$ is also the magnetic moment per unit volume or intensity of magnetization.

$$\vec{H}_m = \vec{M}$$

$\vec{B} = \propto_o \left(\overline{H} + \overline{M}\right)$, it follows that the vector,

$$\vec{M} = \frac{\vec{B}}{\propto_0} - \vec{H} \qquad (2.74)$$

The quantity $\vec{H}$ has the same units as $\vec{M}$. From equation (2.74), the total magnetic field $\vec{B}$ is written a:

$$\vec{B} = \propto_o \vec{H} + \propto_o \vec{M} = \propto_o \left(\vec{H} + \vec{M}\right) \qquad (2.75)$$

Part of the total magnetic field represented by $\vec{H}$ is due to external factors such as the current in the solenoid. The other part represented by $\vec{M}$ is due to the specific nature of the magnetic material that can be influenced by external magnetic fields. *M* is directly proportional to H and can be written as:

$$\vec{M} = \chi \vec{H} \qquad (2.76)$$

The proportionality constant χ is called *magnetic susceptibility* of the material. χ is a dimensionless constant. It is a measure of how a magnetic material can be magnetized. From equations (2.75) and (2.76) the total magnetic flux density becomes:

$$\vec{B} = \propto_o \left(1 + \chi\right) \vec{H} \qquad (2.77)$$

$$\vec{B} = \propto_0 \propto_r \vec{H} \qquad (2.78)$$

where $\propto_r = 1 + \chi$ is a variable dimensionless quantity called the *relative magnetic permeability* of a substance.

The *magnetic permeability* of a substance $\propto$ is related to relative permeability as:

$$\propto = \propto_0 \propto_r = \propto_0 \left(1 + \chi\right) \qquad (2.79)$$

Example 2.23

A coil is tightly wound around an iron cylinder having 2400 turns per metre, calculate:

(a) the magnetic field strength due to magnetization current of magnitude 0.22 A; and

(b) the magnetic susceptibility if the magnetic field in the iron is 1.5 T.

Solution

(a) The applied magnetic field B_0 is given as:

$$B_0 = \propto_0 nI_0$$

$$B_0 = \left(4\pi \times 10^{-7}\,\text{mA}\right)\left(2400\,\text{m}^{-1}\right)\left(0.22\,\text{A}\right)$$

$$= 6.6 \times 10^{-4}\,\text{T}$$

(b) The magnetic χ susceptibility is calculated as:

$$\chi = \left(\frac{\text{Field in the iron}}{\text{Applied field}}\right) - 1 \Rightarrow \chi = \frac{B}{B_0} - 1$$

$$\chi = \frac{1.5\text{T}}{6.6 \times 10^{-4}\,\text{T}} - 1 = 2.3 \times 10^3$$

Example 2.24

Calculate the magnetic field intensity of a magnetized substance in which the magnetization is $9 \times 10^5\,Am^{-1}$ and the magnetic field has magnitude of 5.3 T.

Solution

The magnetic field intensity, H is calculated using the following formula:

$$B = \propto_0 \left(H + M\right) \Rightarrow H = \frac{B}{\propto_0} - M$$

$$H = \left(\frac{5.3\text{T}}{4\pi \times 10^{-7}\,\text{TmA}^{-1}}\right) - 9 \times 10^5\,\text{Am}^{-1}$$

$$= 3.32 \times 10^6\,\text{Am}^{-1}$$

Example 2.25

A cylinder made of soft-iron with cross-section area of $6.0 \times 10^{-4}\,\text{m}^{-2}$ is bent to form a ring of mean diameter of 0.2 m. The ring formed is wound with 2400 turns of conductor carrying a current of 34 A. If the magnetic flux in the soft iron is $9.0 \times 10^{-3}\,Wb$, what is the value of the relative permeability $\propto_r$ of the core?

Solution

Magnetic flux Φ is given by;

$$\Phi = BA = \propto HA \text{ but } H = \frac{NI}{L}$$

$$\therefore \Phi = \frac{\propto NIA}{L} = \frac{\propto_0 \propto_r NIA}{L}$$

It follows that $\propto_r = \dfrac{\Phi L}{\propto_0 NIA}$

Given;

$\Phi = 9.0 \times 10^{-3}\,\text{Wb}$ $\quad N = 2400,\ I = 34\,A,$ $L = 2\pi\left(0.1\right)$ and area $= 6.0 \times 10^{-4}\,m^2$

$$\propto_r = \frac{\left(9.0 \times 10^{-3}\,\text{Wb}\right)\left(2\pi\right)\left(0.1\,\text{m}\right)}{\left(4\pi \times 10^{-7}\,\text{Hm}^{-1}\right)\left(2400\right)\left(34\,\text{A}\right)\left(6.0 \times 10^{-4}\,\text{m}^2\right)}$$

$$= 91.9\,\frac{\text{Wb}}{\text{HA}}$$

Since $1\text{H} = \dfrac{1\,\text{Wb}}{\text{A}}$ then $\propto_r = 91.9$

2.4.3 Classification of magnetic materials

The origin of magnetism and magnetization in terms of electron magnetic dipole moment and susceptibility χ, are used to classify materials as diamagnetic, paramagnetic and ferromagnetic. Paramagnetic and ferromagnetic materials are those whose atoms have a net magnetic moments whereas atoms of diamagnetic materials have zero net magnetic moment. In terms of the susceptibility χ, a material is diamagnetic if χ is negative, paramagnetic if χ is small and positive while ferromagnetic is characterized by large and positive χ. For diamagnetic material, $\propto_r < 1$ and for paramagnetic material, $\propto_r > 1$ while for ferromagnetic material, $\propto_r >> 1$.

(a) Diamagnetic materials

Magnetic properties of matter are due to magnetic dipole moments associated with orbital motions of electrons around the nucleus of an atom. In the absence of external magnetic field, the atoms of diamagnetic substances have no net magnetic moments and hence do not exhibit magnetism. When external magnetic field is applied, currents are induced in the current loops of electrons according to Faraday's laws of electromagnetic induction. According to Lenz's law, electrons having orbital magnetic moment in the same direction as the external magnetic field slow down and those in the opposite direction speed up. Thus, the substance develops a net magnetic moment in direction opposite to that of the applied field and hence repulsion. For diamagnetic substances, $\vec{M}$ is negative and is opposite to $\vec{H}$. The susceptibilities of some materials that exhibit diamagnetism are shown in Table 2.1.

Table 2.1: Magnetic susceptibilities of some substances

Materials	Susceptibility, χ
Silver	-4.2×10^{-6}
Bismuth	-1.66×10^{-5}
Copper	-9.8×10^{-6}
Gold	-3.6×10^{-5}
Lead	-1.7×10^{-5}
Silicon	-1.4×10^{-6}

Diamagnetism is present in all substances, but the effect is weak for most substances hence it is overshadowed by paramagnetism and ferromagnetism.

A certain class of material known as superconductors exhibits perfect diamagnetism. Susceptibility of a superconducting material is negative and hence, using equation (2.76), the material will completely repel external magnetic fields. The phenomenon of perfect diamagnetism in superconductors is called the Meissner effect, after the name of its discoverer (1933). Superconducting materials can be used in a variety of applications including the construction of magnetically levitated superfast trains. Figure 2.37 is an illustration of the Meissner effect shown by a permanent magnet being suspended above a cooled superconductor. Note that superconductors offer zero resistance to the flow of electrical current and they are key to more efficient energy use.

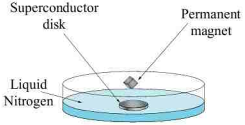


Figure 2.37 Meissner effect

(b) Paramagnetic materials

In paramagnetic materials, individual atoms possess a non-zero net magnetic dipole moment. However due to the random nature of thermal motion of electrons, the material do not exhibit net magnetization. In the presence of an external magnetic field B_0 and at low temperatures, the individual atomic dipole moments align and point in the same direction as the external field Figure 2.38 and hence they are weakly attracted by the field.

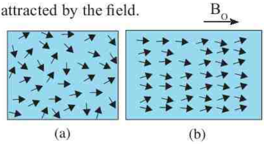


Figure 2.38 (a) Randomly oriented dipole moments (b) Aligned dipole moments

Examples of paramagnetic materials are aluminium, calcium, chromium, lithium, magnesium, nobium, platinum and tungsten. The magnitude of the magnetization vector of a paramagnetic material is inversely proportional to the absolute temperature T and is given by:

$$M = \frac{C}{T} B_0 \qquad (2.80)$$

Since, $M = \chi H$ and $B_0 = \propto_0 H$, equation (2.80) can be written as:

$$\chi = C \frac{\propto_0}{T} \qquad (2.81)$$

Equation (2.81) is known as the *Curie's law,* and the constant C is the Curie's constant. The susceptibility of paramagnetic materials depends inversely on temperature.

(c) Ferromagnetic materials

As in paramagnetic materials, ferromagnetic materials possess non-zero net magnetic dipole moments. In these materials however, the dipole moments interact with one another and spontaneously align themselves into groups each group consisting of magnetic dipole moments in a common direction, even in absence of external magnetic field. The region of space over which the magnetic dipole moments are aligned is called a domain. In the absence of external magnetic field, the domains are randomly oriented as in Figure 2.39 (a), and hence the overall magnetic moment is zero. If the material is placed in an external magnetic field B_o, domains with magnetic moments and domain boundaries aligned with the field as in Figure 2.39 (b); this results in a magnetized material.

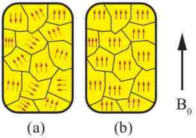


Figure 2.39 Atomic magnetic dipole moments

In some ferromagnetic materials known as magnetically hard materials the magnetization still exist even when the external magnetic field is removed.

Some examples of magnetically hard materials are naturally occurring lodestone and alnico which is an alloy of iron, nickel, aluminium and cobalt. Magnetically hard materials can be used to make permanent magnets. On the other hand, if the magnetization of a ferromagnetic material disappears after the external magnetic field is removed, such materials are known as magnetically soft materials. Examples of magnetically soft materials are iron, cobalt, nickel and mu-metal.

As temperature increases, magnetization decreases and at high temperature the domain structure of a ferromagnetic material disappears, and the material acquires paramagnetic behaviour. The domain disintegration and consequential disappearance of magnetization with temperature is gradual. The transition temperature T_c at which the ferromagnetic material behave as paramagnetic materials is called the Curie temperature. Above the Curie temperature the susceptibility of the material with paramagnetic behaviour is given by:

$$\chi = \frac{C}{T - T_C} \text{ for } \left(T > T_C\right) \qquad (2.82)$$

Below the Curie temperature, the magnetic moments are aligned in ferromagnetic materials. For ferromagnetic materials the value of χ is large and of the order of 10^3 to 10^4 depending on the nature of the magnetic field to which the material is subjected.

(d) Variation of B with H for ferromagnetic materials

Figure 2.40 shows the behaviour of a ferromagnetic material as it goes through magnetization and demagnetization process.

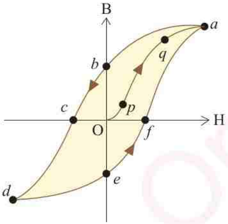


Figure 2.40 Hysteresis loop

Assuming initially the material is not magnetized, in the absence of external magnetizing field H. If magnetizing field, H is applied to a material and increases from zero, the magnetic field B in the material rises from zero at point O. As the magnetizing field increases further there is a shift of domain boundaries and magnetization increases further to p. If the applied field is removed the domain boundaries return to the original positions and the magnetization is again zero. Larger magnetizing fields cause the magnetic axes of entire domains to jump round quite suddenly in succession into alignment with the field and magnetization increases sharply to point q (largely irreversible specimen returns its magnetization if the field is reduced to zero) with greater enough field more or less all domains are in line with the field and saturation occur at point "a". At this point the ferromagnetic has reached saturation. If H is gradually reduced to zero it is found that the curve follows the path ab instead of $aqpO$. At point b, H = 0 but the flux density B in the material has a finite value, called *residual flux density*, also called *retentivity* or *remanence*. Thus, remanence is the magnetism which is left behind when the magnetizing field reduces to zero. On reversing the magnetizing field magnetic field keeps on gradually diminishing and reduced to zero and the curve follows path bc. At point c the magnetic flux density B in the material is zero. The value of H at c needed to wipe out residual magnetism in the material is called *coercive force* or *coercivity*.

The coercive force or coercivity for the material is the magnetizing field which is required to neutralize completely the residual magnetism in the material. As the magnitude of the external magnetic field is increased the magnetic flux density in the material is saturated in the reverse direction depicted by curve cd. If H is reduced to zero at point e the material again retains magnetic field in the opposite direction the remaining part of the loop is obtained by increasing H in the original direction. The curve o,p,q,a,b,c,d,e,f,a is called hysteresis loop. It is clear that B lags behind H and this defines hysteresis as the lagging of B behind H when the magnetic substance is taken through a complete magnetics cycle. The area enclosed by hysteresis loop represents loss in energy appearing as heat in the ferromagnetic material. Thus loss of energy is referred to as hysteresis loss. The shape and size of the hysteresis loop depends on the nature of the material which in turn influence the choice of the ferromagnetic material for a particular application.

2.4.4 Applications of ferromagnetic materials

Magnetically soft materials such as iron are suitable for making electromagnets and cores of transformers Figure 2.41 to increase the magnetic fields produced by current carrying coils. They are specifically used in transformers, moving coil galvanometers and electric bells in which the magnetic field is supposed to vanish as soon as the current is switched off. Ferromagnetic materials are also used in a variety of modern devices such as computer hard disk drives and permanent data storage devices such as thin layer black strip found on the back of ATM cards as shown in Figure 2.42.

Figure 2.41 Transformer

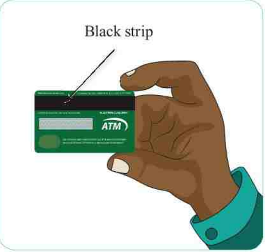


Figure 2.42 The use of ferromagnetic materials

Activity 2.4

Visit an electric power substation to study application of ferromagnetic materials in transformers.

On the other hand, magnetically hard materials, such as steel and other alloys, are for making permanent magnets where large magnetizing fields are required to magnetize them. Once magnetized, the magnetization is retained to a large extent even if the magnetizing field is removed as the rentetivity is large. The magnetization is retained even if the material is exposed to a stray reverse field due to large coercive force.

Exercise 2.4

1. Explain the following terms regarding magnetic properties of materials.
   (a) Magnetic permeability
   (b) Magnetic intensity
   (c) Magnetic susceptibility
   (d) Relative permeability
2. When a material is placed in a toroid carrying a current, the resultant magnetic field B is a sum of two terms: $B = B_o + B_m$. Explain the origin of the two terms.
3. Distinguish among the following materials with reference to magnetic behaviour.
   (a) Ferromagnetic materials
   (b) Diamagnetic materials
   (c) Paramagnetic materials
4. Briefly explain the following terms:
   (a) Currie's temperature
   (b) Hysteresis loop
5. A diamagnetic material is brought close to a permanent magnet. What happens to the material?
6. A current of 1.5 A flows through the windings of a large, thin toroid with 200 turns per metre. If the core of the toroid is filled with iron for which the magnetic susceptibility is 3×10^3, what is the value of the magnetic field?
7. Given that the magnetic dipole moment of an iron atom is $2.1\times10^{-23}\,\text{Am}^2$, calculate:
   (a) the magnetic dipole moment of a domain consisting of 10^{19} iron atoms.; and
   (b) the current flowing through a single circular loop of wire of diameter 1.0 cm that will produce magnetic dipole moment in part (a).
8. Explain three practical applications of ferromagnetic materials.
9. Briefly discuss only five of the wide range of magnetic properties exhibited by ferromagnetic materials.

2.5 Magnetic field of the earth

The Earth has a magnetic field produced mainly from its interior and magnetosphere which forms a protecting shield around the planet. This field is often called the *geomagnetic field*. It is common experience that a compass needle is used for navigation. This device utilizes the earth's magnet fields which is aligned with vertical plane that passes through the poles of the earth. Nowadays, the use of compass needle for navigation has been almost replaced by more accurate navigation systems including the Global Positioning System (GPS). The major benefit of the earth's magnetism is that it prevents ionizing energetic charged particles emitted by the sun from reaching the earth's atmosphere. If these particles enter the atmosphere they would ionize the gas molecules making them unsuitable for supporting life. The protection is achieved by deflection of incident energetic charged particles away from the earth by the geomagnetic field.

2.5.1 Origin of the earth's magnetic field

The origin of the geomagnetic field is still not well understood although there are several theories that try to explain it. The most acceptable theory behind the origin of the geomagnetic field is popularly known as the *dynamo effect*. The earth's core is composed mainly of solid iron in the inner core and liquid iron in the outer core. In this theory the magnetic field is thought to arise due to electric currents produced by convective motion of metallic fluids in the outer core of the earth. The complex motion of the metallic fluids is driven by convection and the rotation of the Earth.

Other sources of the geomagnetic field include magnetized rocks on the earth's crust, electric currents in the ionosphere and magnetosphere, as well as currents induced in different parts of the earth (crust, mantle and oceans) by varying magnetic fields. There is also a contribution from the solar wind.

2.5.2 Structure of the earth's magnetic field

The geomagnetic field lines presented in Figure 2.43 have a close resemblance to those of a bar magnet placed along an axis that passes through the center of the earth. This axis is currently tilted at an angle of about 11.3º from the axis of rotation of the earth. As indicated by the compass needle, the magnetic poles near the geographical North and South poles of the earth are called the South and North magnetic pole, respectively. The naming of magnetic poles follows the actual North and South poles of earth's magnetic dipole represented by a hypothetical permanent bar magnet shown in Figure 2.43.

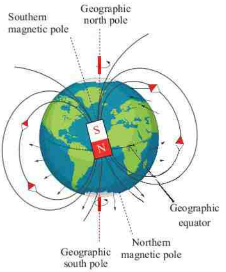


Figure 2.43 Earth's magnetic field

2.5.3 Components of the earth's magnetic field

The geomagnetic field $\vec{B}_E$ being a vector field has direction and magnitude at every point in space and inside the earth. At point P on the Earth's surface, this vector can be described in terms of three components: The north, east and vertical components denoted by X, Y and Z as shown in Figure 2.44 (a).

The geomagnetic field $\vec{B}_E$ is often described in terms of its magnitude B_E, the inclination (angle of *dip*), *I*, defined as the angle to the vector $\vec{B}_E$ makes with the horizontal plane and the declination, *D* defined as the angle between the component (x) and the horizontal component of $\vec{B}_E$ denoted by H. The magnetic field can also be specified in terms of the horizontal and vertical components H and Z and the declination, D Figure 2.44 (b). The relationship between these components is given by the following set of equations:

$$B_E = \sqrt{X^2 + Y^2 + Z^2}$$

$$H = \sqrt{X^2 + Y^2}$$

$$I = \arctan\left(\frac{Z}{H}\right)$$

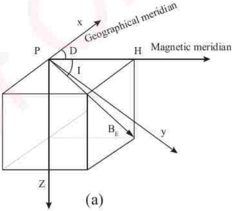
$$D = \arctan\left(\frac{Y}{X}\right)$$

$$B_E = \sqrt{H^2 + Z^2} \qquad (2.83)$$

In Cartesian coordinates, the components are:

$$X = H\cos D$$
$$Y = H\sin D$$
$$Z = B_E \sin I$$

where *"I"* and *"D"* are angles in degrees.



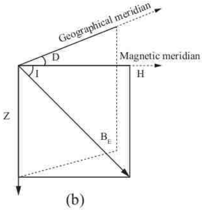


Figure 2.44 (a) Components of the Earth's magnetic field
(b) Angle of declination

Example 2.26

The horizontal component of the earth's magnetic field, H, at a certain location is 4×10^{-5} T If the angle of dip is 60^0 calculate the vertical component and total intensity of the earth's field.

Solution

The horizontal component, $H = 4\times10^{-5}$ T and the angle of dip is 60^0,

$$I = \arctan\left(\frac{Z}{H}\right) \Rightarrow Z = H\tan(I)$$

$$Z = 4\times10^{-5}\,\text{T}\times\tan(60^\circ) = 6.93\times10^{-5}\,\text{T}$$

For the total intensity of the earth's field, B_E, gives;

$$H = B_E\cos(I) \Rightarrow B_E = \frac{H}{\cos(I)}$$

$$B_E = \frac{4\times10^{-5}\,\text{T}}{\cos 60^0} = 8\times10^{-5}\,\text{T}$$

2.5.4 Variations of the earth's magnetic field

Measurements have shown that there are two classes of geomagnetic field variations namely spatial and temporal variations. The time variations of the geomagnetic field can be short term or long term variations that can be periodic or random, taking place in time intervals ranging from seconds to millions of years. The variations are in form of changes in direction and/or strength of the field. Apart from duration, the difference between short term variations and long term variations is that long term variations are caused by the dynamics of the Earth's interior while the short term variations are caused by external factors such as rotational motions of the moon, earth and sun. The long term variations are on a scale of 5 years or more and are called *secular variations*.

Paleomagnetism is a study of the record of the Earth's magnetic field in rocks, sediment, or archeological materials and is the science used to determine the direction and intensity of the Earth's magnetism in the past. The most important contribution of this study was the discovery of the *magnetic field reversal* which is the change in the declination, D by 180^o which leads in the exchange of position between the north and south magnetic poles as shown in Figure 2.45. Analysis of remnant magnetization in volcanic and sedimentary rocks have revealed that the last magnetic field reversal occurred around 780,000 years ago and the transition was made in the time interval ranging from 5,000 to 10,000 years. In addition, the magnetic north pole could also experience a slow shift in position relative to the geographic north in what is known as *polar wandering*.

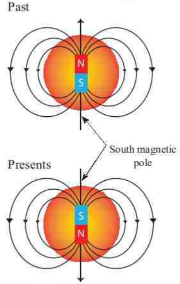


Figure 2.45 Reversal of the magnetic north pole

Variations whose duration do not exceeds a year are called short term variations. The short term earth's magnetic field variations can be intense and are mainly produced by currents in the magnetosphere and ionosphere.

Exercise 2.5

1. The earth's magnetic inclination angle in Dar es Salaam is about 35°. Would you expect a smaller or greater inclination angle in South Africa? Explain.
2. Suppose a magnetic compass is located right on the geomagnetic south pole, in which direction would you expect the compass to point? Explain your answer.
3. List three components used to specify the earth's magnetic field.
4. Describe two theories behind the origin of the earth's magnetic field.
5. The earth's magnetic field varies from point to point in space. Does it also change with time? If so, on what time scale does it change appreciably?
6. What evidence suggests that the earth has undergone magnetic pole reversal?

Revision Exercise

1. Discuss the similarities and differences between the electrical force and the magnetic force on a charged particle.
2. An electron moving along the positive y axis perpendicular to a magnetic field experiences a magnetic deflection in the negative x direction. What is the direction of the magnetic field?
3. If a charged particle moves in a straight line through some region of space, can you say that the magnetic field in that region is necessarily zero? Explain.
4. A velocity selector in a Bainbridge mass spectrometer uses a magnetic field of strength 0.16 T. Calculate:
   (a) the electric field strength that is needed to select a speed of $3.5\times10^{6}\text{ m/s}$.
   (b) the *p.d.* between the plates if they are separated by 1.5 cm?
5. In the Bainbridge mass spectrometer a beam of singly ionized neon atoms moving with the same speed of $5.5\times10^{6}\text{ m/s}$ is introduced into the velocity selector region. When the beam exits the velocity selector, it enters a uniform magnetic field region and follows a circular path of radius 0.2 m. Use this information to answer the following questions:
   (a) explain why the particles follow a circular path;
   (b) state the rule which can be used to predict the direction of the

magnetic force as the particles follows a circular path;

(c) calculate the magnitude of the force on each neon ion of mass of 3.32×10^{-26} kg;

(d) calculate the magnetic flux density into which the ions enter after exiting the velocity selector;

(e) if the velocity selector only allows ions travelling at speed of 5.5×10^{6} m/s to pass without deflection;

(i) calculate the electric field in this region given the magnetic flux density in the velocity selector is 3.0×10^{-2} T; and

(ii) using the results obtained in (e) (i), calculate the potential difference between the parallel plates in the velocity selector if the plate separation distance is 0.15 m.

6. A circular coil of radius 18 cm is wound with twenty turns and carries a current of 4.3 A. Calculate the maximum torque if the coil is placed in a uniform magnetic field strength of 5 T.

7. Why is it possible for a nearby magnet to distort a television picture produced by a cathode ray tube?

8. A circular coil with 100 turns has a diameter of 0.04 m. Calculate:
   (a) Current through the coil to provide a magnetic dipole moment of $4\,\text{Am}^2$.
   (b) Maximum torque that the coil will experience in a uniform magnetic field of strength 0.1 T.
   (c) If the angle between the magnetic dipole moment μ and magnetic field of strength B is 45°, what is the magnitude of the torque on the coil?

9. Explain why a current-carrying conductor experiences a force when placed in an external magnetic field.

10. Describe how should a current-carrying conductor be placed in a uniform magnetic field so that it experiences no magnetic force.

11. How can the motion of a charged particle be used to distinguish between a magnetic and an electric field?

12. Describe the effects of a magnetic field on a moving charge.

13. A proton moves at 6×10^{7}m/s perpendicular to a magnetic field. The field causes the proton to travel in a circular path of radius 0.8 m. Calculate the magnetic field strength.

14. Does changing the direction of a magnetic field necessarily mean a change in the force on a moving charged particle? Explain your answer.

15. A particle of charge q and mass m moving with a speed v enters a region of uniform magnetic field of magnitude B at an angle of θ to the magnetic field. In the magnetic field the particle describes a helical path with radius R and distance between loops (pitch) P. Derive expressions for R and P.

16. A particle with a charge twice that of an electron moves through a uniform magnetic field of strength 0.79 T perpendicular to the direction of its velocity. Determine the mass

of the particle if it has a cyclotron frequency of 1.6 KHz.

17. What is a cyclotron? Discuss the construction and operation of a cyclotron.
18. Compare Biot-Savart law for magnetic fields with Coulomb's law for electrostatic fields.
19. Using a well labelled diagram, illustrate the Biot-Savart law.
20. For what conditions is Ampere's law preferable over the Biot-Savart law?
21. Identical currents are carried in two circular loops; however, one loop has twice the diameter of the other loop. Compare the magnetic fields created by the loops at the center of each loop.
22. How would you orient two long, straight, current carrying wires so that there is no net magnetic force between them?
23. Explain why $\bar{B}=0$ inside a long, hollow copper pipe that is carrying an electric current parallel to the axis.
24. A current of 5 A flows in a circular loop with 120 turns of wire and radius 50 cm. Find the magnetic flux through the loop.

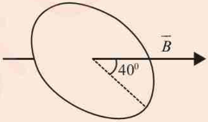


Figure 2.46 Circular loop

25. A solenoid has 1500 turns/metre of wire. It is 25 cm long, and 6 cm in diameter. What is the magnetic flux through the solenoid when a current of 2.5 A flows through the wire?
26. Is Ampère's law valid for all closed paths? Explain.
27. A wheel with 8 metallic spokes each 50 cm long is rotated with a speed of 120 rev/min in a plane normal to the horizontal component of the earth' magnetic field. The earth's magnetic field at the plane is $0.5\times10^{-5}\,\text{T}$ and the angle of dip is 60°. Calculate the *e.m.f.* induced between the axle and the rim of the wheel.
28. At a place, the horizontal component of earth's magnetic field is B and angle of dip is 60°. Calculate the horizontal component of the earth's magnetic field at the equator.
29. Describe an experiment to demonstrate electromagnetic induction.
30. Describe, with the aid of a diagram how a magnet and a solenoid can be used to produce electricity.
31. (a) Explain using a labelled diagram how you can use electromagnetic induction to construct a seismometer.
    (b) Use your diagram to explain how a seismic record would look like as a result of up-down movement of the earth on which the instrument is placed.

32. Two circular loops A and B have their planes parallel to each other, as shown in Figure 2.47.

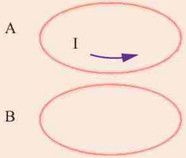


Figure 2.47 Circular loops

Loop A has a current moving in the counter clockwise direction, when viewed from above.

(a) If the current in loop A decreases with time, what is the direction of the induced current in loop B? Will the two loops attract or repel each other? Explain.

(b) If the current in loop A increases with time, what is the direction of the induced current in loop B? Will the two loops attract or repel each other? Explain.

33. Electromagnetic braking can be achieved by applying a strong magnetic field to a spinning metal disk attached to a shaft.

(a) How can a magnetic field slow the spinning of a disk?

(b) Would the brakes work if the disk was made of plastic instead of metal?

34. A metal rod is forced to move in a magnetic field of strength 0.35 T with constant velocity along two parallel metal rails.

(a) If the rails are separated by 25 cm and the speed of the rod is 55 cm/s, calculate the induced *(e.m.f.)* .

(b) If the rod has a resistance of 18 Ω and the rails and connector have negligible resistance, what is the current in the rod?

(c) What is the rate at which energy is being transferred to thermal energy?

35. Derive an expression for induced e.m.f. as a result of a varying magnetic field in the following aspects:

(a) Rectangular loop near a wire.

(b) Sliding rod.

(c) Moving loop.

36. (a) Describe one aspects by which an e.m.f. can be induced in a conductor or coil.

(b) What factors determine the magnitude and direction of the induced e.m.f. in (a)?

(c) Describe an experimental set-up that can be used to investigate the factors on which the induced e.m.f. depends.

37. Consider a coil of 400 turns of cross sectional area 30 cm^2 connected in a circuit of resistance 200 Ω. What will be the maximum charge induced in the coil when it is inserted in a magnetic field of flux density 2.5×10^{-3} T to produce a maximum change of flux? Ensure that all the units of quantities used in the calculations are shown and appropriate conversion factors are used to obtain coulomb as unit of charge as the answer.

38. With the help of a sketch diagram explain how a metal disc mounted on a wheel axel could use magnetic induction of a coil connected in series with a 24 V battery, rheostat and brake pedal to increase braking effectiveness of the force exerted on the brake pedal by the foot of the driver of a heavy vehicle. Hint: The brake should have a lever that reduces resistance when the brake is applied.
39. A bar magnet falls vertically with N above the S pole towards the center of a horizontal loop of a conductor. How is the acceleration of the bar magnet affected by the gravity?
40. Suppose the poles of the magnet were reversed and the experiment in problem 39 was repeated. What will happen to the acceleration of the magnet due to gravity?
41. The coupling coefficient of two coils is 0.7. What does this value mean?
42. A copper ring is suspended in a vertical plane by a piece of thread. What will happen to the ring if it is approached by a bar magnet travelling horizontally towards its center?
43. A coil of inductance $L = 0.50$ H is connected to a battery of 12.0 V. By using appropriate quantities with units, show that the growth rate of current in this set-up is 24.0 As^{-1}.
44. Describe the Lenz's law?
45. How does the inductance of a coil change when an iron rod is placed inside the coil?
46. If the self-inductance of an air core inductor increases from 0.001mH to 100 mH by replacing the core with iron, what is the value of the relative permeability $\propto_r$ for iron?
47. A spherical conducting shell is placed in a time-varying magnetic field. Is there an induced current along the equator? Justify your answer.
48. A rectangular loop moves across a uniform magnetic field but the induced current is zero. How is this possible?
49. Define magnetic flux and magnetic flux linkage.
50. Explain the term electromagnetic induction.
51. State Faraday's laws of electromagnetic induction. Describe an experiment to demonstrate Faraday's laws.
52. Discuss the factors that influence the induced *e.m.f.* in a closed loop of a wire carrying a current.
53. A coil of N turns is connected in a circuit of resistance R. If the flux linking each turn is changing, show that the total charge Q induced when the flux changes from Φ_0 to Φ_1 is given by $Q = \dfrac{N(\Phi_1 - \Phi_0)}{R}$.
54. An electric generator consists of 120 turns of wire formed into a rectangular loop of area 1440 cm^2, placed entirely in a uniform magnetic field of strength 4T. Calculate the induced *e.m.f.* when the loop is spun at 3000 rev/min about an axis perpendicular to the magnetic field.

55. What factors contribute to the total magnetic dipole moment of a hydrogen atom?
56. Why is the susceptibility of a diamagnetic material is negative?
57. Discuss the difference among diamagnetic, paramagnetic, and ferromagnetic materials.
58. What is the difference between hard and soft ferromagnetic materials?
59. Should the surface of a computer disk be made from a hard or a soft ferromagnetic material?
60. Given only a strong magnet and a screw driver, how would you first magnetize and then demagnetize the screw driver?
61. Explain why it is desirable to use hard ferromagnetic materials to make permanent magnets.
62. If you cut a bar magnet into two pieces, will you end up with one magnet with an isolated north pole and another magnet with an isolated south pole? Explain your answer.
63. Explain why some atoms have permanent magnetic dipole moments and others do not.
64. A soft iron ring has a mean diameter of 0.2 m and cross-section area of $6.0\times10^{-4}\,\text{m}^2$. If the ring is uniformly wound with 2400 turns of a conductor carrying 2 A, what would be the relative permeability of the iron if the magnetic flux in the iron is $8.0\times10^{-3}\,\text{Wb}$?
65. The total flux density B in the core of a coil carrying a current I is given by $B = \propto_0 \left(H + M\right)$ where H is the magnetizing field intensity due to the current and M is the magnetization produced in the material. Use the relationship to show how relative permeability $\propto_r$ and the susceptibility χ of the material are related.

Chapter Three

Electronics

Introduction

On a daily basis, people interact with different types of devices such as phones, radio, watches, cameras, televisions, decoders, video recorders, security lamps, alarm systems and other display systems. These devices are used in information technologies, telecommunication, industries, and they use electronics in their operations. The study of electronics provides physicists with the knowledge and expertise of manufacturing and maintaining electronic devices. In this chapter, you will learn about the fundamental concepts of electronics, such as theory of semiconductors, energy bands in solids as well as basic electronic components like diodes, transistors, logic gates and operational amplifiers. You will also be introduced to telecommunications as an application of electronics.

3.1 The band theory of solids

The electrical properties of a solid can be described based on the band theory of solids. For a single, isolated atom as shown in Figure 3.1 (a), and the electrons in any orbit (shell) associated with a definite energy levels as shown Figure 3.1 (b). Conversely, in solids, atoms are closely packed together, therefore, affecting the energy levels of the outmost shell electrons of the neighbouring atoms. The results is that the electron in any orbit of such an atom can have a wide range of energy levels to occupy. The wide range of energy levels that can be occupied by an electron in the solid is known as *energy bands*.

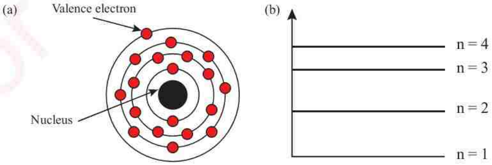


Figure 3.1 Atomic structure and energy levels

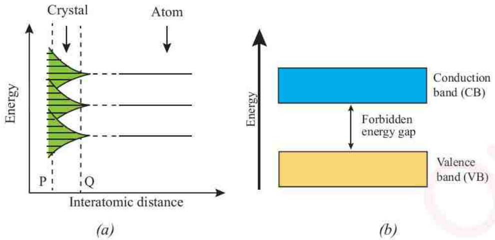


Figure 3.2 Energy bands in solids

The energy levels of isolated atoms are transformed into equivalent energy bands when the atom is in a solid as shown in Figure 3.2 (a). There is no substantial change in the energy levels of electrons in the inner orbits of an atom but there is substantial change of energy levels of electrons in the outer orbit (valence electrons). This is because valence electrons are shared by more than one atom in a crystal. Therefore, valence electrons can be considered to be in either of two energy bands i.e. valence band or conduction band.

Free charge carriers are responsible for electrical conduction in solids. The free electron theory suggests that solids which do not possess free electrons are insulators and those which possess are conductors. It is assumed that the free electrons can have any energy. The band theory permits the free electrons to have only certain ranges of energies. For an isolated atom as in a gas, there is a characteristic set of well-defined energy levels that electrons can occupy. If the atoms are far apart, their orbital interactions are negligible. However, if the atoms are brought uniformly close to each other, the electric field due to their charges overlap and mutual interaction occurs, resulting to a range of the energy bands.

3.1.1 Valence band

When the orbitals (shells) interact for the identical atoms, half of the energy levels are lowered in energy, and another half are raised in energy in respect to the sum of the energies of the valence electron orbitals. Valence electrons fill or occupy the lowered energy levels in pairs of two according to Pauli's exclusion principle, leaving behind the un-filled or un-occupied energy levels. The range of energy levels filled with valence electrons is called valence band as depicted in Figure 3.2 (b). Always valence band is filled with electrons.

3.1.2 Conduction band

The upper most bands are called conduction band as illustrated in Figure 3.2 (b). The valence electrons can be excited from the valence band to the upper band by application of external energy such as heat or electrical energy. At room temperature, for example, some valence electrons can be excited from the valence band to conduction band where they can move freely in the crystal. These free electrons are responsible for conduction of electricity in materials. The free electrons are sometimes known as conduction electrons. The conduction band, therefore, is the range of un-filled or partially filled energy levels, where electrons have great mobility within the crystal.

3.1.3 Forbidden energy gap

The gap or separation between the valence band and conduction band as depicted by two direction line in Figure 3.2 (b) is called the forbidden energy gap or energy gap. Band gap, therefore, is the difference in energy between the highest energy state in the valence band and the lowest energy state in the conduction band. No electron can stay in the forbidden energy gap, as there is no allowed energy state. For an electron to move from the valence band to the conduction band, an external energy (e.g. thermal electrical energy) at least equal to the forbidden energy gap must be applied. The size of the forbidden gap gives a measure of bondage of the valence electrons to the atoms. The greater the size of energy gap, the stronger the valence electrons are attracted by the nucleus. In such a case, more external energy is required to excite an electron from the valence band into the conduction band.

3.1.4 Classification of solids

On the basis of the band theory of solids, there are three types of solid materials namely conductors, semiconductors and insulators.

(a) Conductors

These are solids such as iron, copper and aluminium through which electric current can easily flow. In terms of the band theory, conductors are solids whose valence band (VB) and conduction band (CB) overlap i.e. no forbidden gap (FG) as depicted in Figure 3.3 (a). At room temperature, the valence electrons gain some energy and move to the conduction band. A conduction band contains a significant number of free electrons which are responsible for electrical conduction. Therefore, even a small potential difference applied across a conductor at a room temperature will constitute a large electric current.

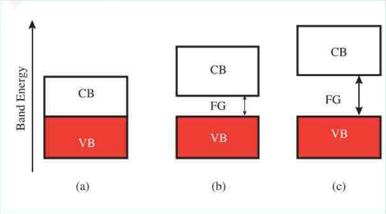


Figure 3.3 Energy band diagrams for (a) conductor, (b) semi-conductor and (c) insulator

(b) Semiconductors

These are solid materials whose electrical conductivities lie between those of conductors and insulators. It is well-known that the resistivity of a conductor is about $10^{-8}\ \Omega m$, for semiconductor is about $10^{-1}\ \Omega m$, while that of an insulator is about $10^{4}\ \Omega m$. Elemental semiconductors mostly constitute elements of Group IV in the Periodic Table, and the commonly used ones are silicon (Si) and germanium (Ge). In terms of band energy, semiconductors lie between conductors and insulators as shown in Figure 3.3 (b). The forbidden energy gaps are very small to about 0.3 eV in germanium and 0.7 eV in silicon. At room temperature, the valence band is almost filled with electrons and the conduction band is almost empty. As the temperature increases, more valence electrons cross the forbidden gap to the conduction band, thereby increasing the electrical conductivity.

(c) Insulator

Insulators are materials such as wood, glass and mica through which electric current can hardly flow. In terms of the band theory, insulators are materials whose valence band and conduction bands are separated by a wider forbidden gap as shown in Figure 3.3 (c). The valence band of an insulator may be partially or completely filled with electrons while its conduction band is an empty. At room temperature, valence electrons cannot cross the forbidden energy gap of an insulator because the energy provided is insufficient to excite valence electrons from the valence band to the conduction band. When the temperature is highly raised to a certain value, some of the valence electrons may acquire enough energy (*5* eV to 15 eV) to cross the energy gap and reach the conduction band, hence the insulator becomes a conductor.

3.1.5 Fermi level

In the band theory of solids, the energy level that has half probability of being occupied at any given time is called Fermi level, as named after a physicist Enrico Fermi (1901-1954) who first proposed it. The Fermi level is also named as the electrochemical potential of a system, that is a work done required to add an electron in a system. At room temperature, the number of holes in the valence band equal the number of electrons in the conduction band for pure materials. Therefore, Fermi level (half-probability of occupancy) for a conductor sits at the middle of the overlapping bands, whereas for a pure semiconductor and insulator the Fermi level is at the middle of the forbidden energy gap as illustrated in Figure 3.4.

Fermi level can be used to determine the number of electrons per unit volume and velocity of electrons in a solid, which is important in understanding the flow of electrons in solids. It helps to explain why electrons do not contribute significantly to the specific heat capacity of solids at ordinary temperatures while they are dominant contributors to thermal and electrical conductivities. It is important to note that, Fermi level is different from Fermi energy. Fermi energy, E_F is the energy of the highest occupied state at absolute zero temperature (-273.15 °C or K). Fermi energy is defined only at absolute zero temperature whereas Fermi level is defined for any given temperature including absolute zero.

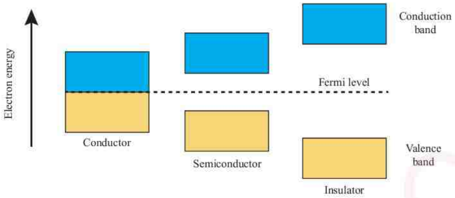


Figure 3.4 Fermi level of solid materials at room temperature

Activity 3.1

By the use of a the Periodic Table and any other source of information, identify elements which are good conductors, semiconductors and insulators. In each type arrange the elements according to their conductivity.

3.1.6 Effect of temperature on the electrical conductivity of solids

The electrical conductivity of solid materials changes with temperature. In case of conductors, conductivity decreases with increase in temperature. This is because as the temperature increases the amplitudes of vibration of the atoms in the conductor increases, and produces hindrance of charge transfer by the flow of electrons, thus decreased conductivity. For the semiconductors, an increase in temperature results to an increase in thermal energy of the valence electrons which enables more of them to break the valence bonds and become free electrons. More electron-hole pairs are thermally generated and act as carriers of current, thus increase the conductivity of semiconductors. At absolute zero all semiconductors behave as insulators.

Comparatively, the temperature change in an insulator, in most cases does not affect the electrical conductivity, as no electron jump from valence band to the conduction band. It is important to note that a superconductor has perfect conductivity at temperatures approaching absolute zero (0 K).

Activity 3.2

Design an electric circuit to study the variation of current flow through various metallic conductors of the same dimensions.

Exercise 3.1

1. Describe the formation of energy bands in solids.
2. Explain the significance of the Fermi level in solids.
3. Explain the difference between conductors, semiconductors and insulators with the help of an energy band diagram.

4. Explain the changes in electrical conductivity of conductors and semiconductors as temperature increases.
5. Use the band theory of solids to explain why insulators do not conduct electric current.

3.2 Classification of semiconductors

As discussed previously, some of Group IV elements of the Periodic Table are semiconducting materials. Silicon (Si) and germanium (Ge) are the best known semiconductors; they are used to make common electronic devices such as transistors and diodes. Cadmium sulphide, lead sulphide, and gallium arsenide are examples of semiconductor compound materials. Semiconductors are categorized as intrinsic or extrinsic semiconductors depending on purity and conductivity.

3.2.1 Intrinsic semiconductors

A pure semiconductor is often called an intrinsic semiconductor. This kind of semiconductor cannot conduct electricity unless the electrons are excited to the conduction band from valence band by external energy (thermal or electrical). An electron excited to the conduction band leaves behind a vacancy termed as a hole in the valence band as shown in Figure 3.5. The holes are positive charge carriers while electrons are negative charge carriers. In intrinsic semiconductor, the number of holes is equal to that of electrons in their respective energy bands. Application of an electric field in materials causes the adjacent electrons in the valence band to occupy the holes, but in doing so they generate other holes. The valence electrons, and holes move in the opposite direction and contribute to a small current in a pure semiconductor. It is said to exhibit intrinsic conduction i.e. charge carriers have their origin inside the material.

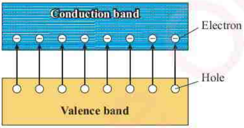


Figure 3.5 Formation of holes and electrons in an intrinsic semiconductors

3.2.2 Extrinsic semiconductors

An intrinsic semiconductor has relatively few charge carriers, and to increase its conductivity, a tiny amount of impurity is introduced into it. The conductivity and other electronic properties of a pure semiconductor can be changed in a controlled manner. The process of adding controlled impurities to a semiconductor is known as doping. The amount of impurities or dopants added to an intrinsic semiconductor changes its level of conductivity. The impurity atoms added may be either pentavalent or trivalent elements. The doped semiconductors are termed as extrinsic semiconductors. This is because the added impurity to the materials add more charge carriers. There are two types of extrinsic semiconductors namely negative type *(n-type)* and positive type *(p-type)* semiconductors, where n and p stands for negative and positive charges respectively.

(a) n-type semiconductors

An intrinsic semiconductor such as *Si* can be converted into n-type semiconductor by *doping* it with pentavalent (Group V) elements in the Periodic Table such as phosphorus (P) and arsenic (As). In case one phosphorous atom is added to the Si, the phosphorus atom settles in the lattice site with four of its electrons forming a covalent bond with Si atoms while the fifth electron becomes free to roam through the crystal as shown in Figure 3.6. The impurity atom provides one free electron resulting to a large increase in the number of electron carriers. Therefore electrons become the majority carriers and holes are minority carriers. The pentavalent atoms are known as donors since they donate one free electron from each atom.

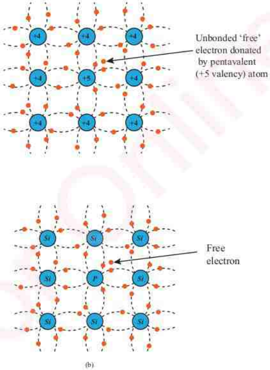


Figure 3.6 n-type semiconductor

In band energy level diagram, Figure 3.7 indicates the pentavalent impurity atom creates electron in silicon, a donor energy level just below the conduction band.

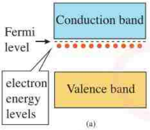


Figure 3.7 The energy level for donor atoms

(b) p-type semiconductors

An intrinsic semiconductor such as germanium *(Ge)* can be converted into a p-type semiconductor by doping it with trivalent (Group III) elements in the Periodic Table such as boron (B), indium (In) and gallium (Ga). In case one gallium atom is added to the Ge, each gallium has only three electrons and therefore can form covalent bonds with only three of its four germanium neighbours as illustrated in Figure 3.8 (a). It needs very little energy for an electron in a nearby Ge - Ge bond to move through and occupy the vacancy in the gallium. Under normal conditions lattice vibrations provide this energy, creating a hole in one of the germanium atom because trivalent atoms accept electrons from the crystal for conduction, and in this way they are referred to as *acceptors*. The majority carriers in a p-type semiconductor are holes while electrons are minority carriers. In band energy level diagram, the trivalent impurity atom creates a hole in the germanium atom, and an accepter energy level is just above the valence band as shown in Figure 3.8 (b).

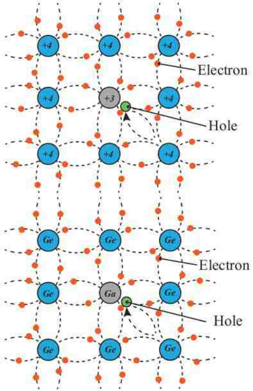
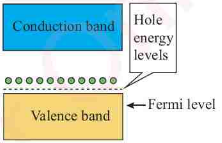


Figure 3.8(a) p- type semiconductor

In band energy level diagram, the trivalent impurity atom creates a hole in the germanium atom, and an accepter energy level is just above the valence band as shown in Figure 3.8 (b).



(b)

Figure 3.8(b) The energy level for acceptor atoms

3.2.3 The p-n junction

When a p-type material and an n-type material are joined side by side by special techniques, a p-n junction is formed as shown in Figure 3.9 (a). In isolation, p-type and n-type materials contain an imbalance of charge carriers as already noted. In contact, holes diffuse from the p-type towards the n-type and electrons from n-type to p-type due to carrier concentration gradient across the junction, as shown in Figure 3.9 (b). The diffusion of holes and electrons across the boundary sets up a potential barrier which prevents further diffusion of charge carriers. The size of the potential barrier depends on the types of material, amount of doping and temperature. The potential barrier of the p-n junction was found experimentally to range between 0.2 V - 0.3 V for germanium and 0.6 V - 0.7 V for Silicon. The diffusion process only occurs over a very thin region known as a *depletion layer* as represented in Figure 3.9 (c).

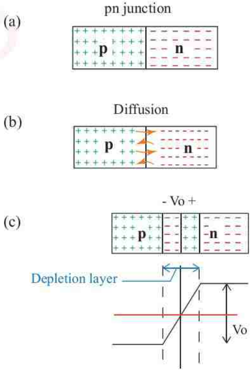


Figure 3.9 Diffusion of charge carriers in a p-n junction

(a) Biasing of a p-n junction

Biasing is the application of an external voltage of proper magnitude and polarity across a p-n junction. There are two types of biasing of the p-n junction, namely, *forward biasing* and *reverse biasing*. The p-n junction is said to be *forward biased* when the p-side is connected to a positive terminal of a battery and the n-side is connected to the negative terminal as shown in Figure 3.10 (a).

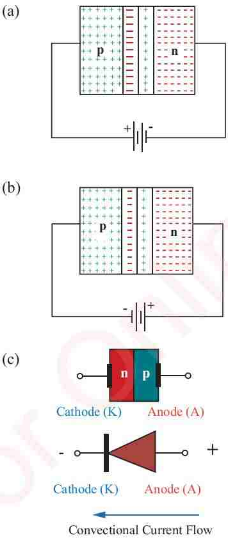


Figure 3.10 Biasing of the p-n junction

When the connection is reversed as illustrated in Figure 3.10 (b), the p-n junction is said to be *reverse biased*. In reverse bias connection, very small current flow due to minority carriers whereas, much current flows in the forward bias connection due to majority carriers. Since, significant current flows only when a p-n junction is forward biased, p-n junction forms a semiconductor diode. The exceptional property of a diode to conduct current in one direction only allows it to be used as a rectifier. The structure and the circuit symbol for a semiconductor diode with a conventional current flow is portrayed in Figure 3.10 (c).

(b) I-V curve characteristic of a p-n junction

The I -V curve characteristics of a p-n junction is the relation between the voltage applied (V) across a junction and the current (I) flowing through it. The characteristic curve of forward biased appear quite different from that of reverse biased as shown in Figure 3.11. When a small external voltage is applied across the p- n junction in forward biased mode, a very small current starts flowing in the p-n junction. As the external voltage increases, the width of depletion region becomes narrow, allowing more charge carriers to cross the junction. A point is reached when the depletion layer is completely eliminated and the current increases rapidly as shown on the right hand side of the vertical axis of Figure 3.11. The minimum voltage at which the depletion layer is completely eliminated is called "*knee*" voltage. For silicon, the knee voltage ranges from 0.6 V to 0.7 V while for germanium, it ranges from 0.2 V to 0.3 V.

In reverse bias condition, a positive terminal of voltage source is applied to the n-type material and a negative terminal of voltage source is applied to the p-type material. The electrons in the n-type material are attracted towards the positive electrode and away from the junction, while the holes in the p-type end are also attracted away from the junction towards the negative electrode. As the reverse voltage increases, the width of the depletion layer becomes wider and hence, practically no current flow through the circuit. However, there is a very small current that flows in reverse bias due to minority charge carriers. The I-V curve for this case is depicted in Figure 3.11 in the left hand side of the vertical axis.

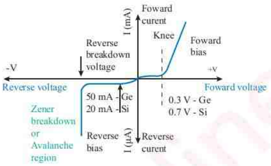


Figure 3.11 I - V Curve characteristic of a p-n junction

As the reverse voltage increases further, the electric field **E**, across the p-n junction becomes strong enough, causing bonds that hold atoms together to break. This breakdown results into release of electrons, which gives rise to a sharp increase of reverse current as shows in Figure 3.11. The voltage at which the breakdown occurs is called breakdown voltage. There are two distinct processes by which breakdown may occur, namely *Zener* breakdown and *avalanche* breakdown.

Activity 3.3

Connect a series circuit using a 6 V battery, variable resistor, milliammeter, switch, diode, and a low voltage bulb.

(i) Switch on the current and adjust the variable resistor from high to reasonable current and note the brightness of the bulb.

(ii) Reverse the terminals of the diode and observe what will happen to the bulb.

(iii) Replace the milliameter with a very sensitive ammeter, note down the values of current while adjusting the variable resistor.

(iv) Write down your observations.

3.2.4 Zener breakdown

A Zener diode is a unique type of a diode that is designed to operate in reverse mode when the Zener voltage, V_z, is applied. This diode is named after Clarence Melvin Zener (1905 - 1993) who discovered the Zener effect. The circuit symbol of Zener diode is shown in Figure 3.12 (a).

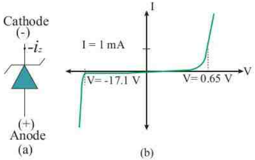


Figure 3.12 A symbol and I-V curve characteristic of Zener diode

A Zener diode is a highly doped p-n junction, so that the depletion region (p-n junction) becomes very thin. When a reverse biased voltage equal to the Zener voltage, V_z, is applied on a Zener diode, it creates an extremely strong electric field, E, across the p-n junction. The strong electric field breaks the covalent bonds producing a large number of electron-hole pairs appearing as minority charge carriers. Note that, when the reverse voltage is applied, the majority charge carriers move away from the junction, while the minority charge carriers move toward the junction.

The depletion layer, normally blocks the minority carriers from crossing the p-n junction. However the Zener diode is designed such that the depletion layer is extremely narrow as a result, the minority charge carriers "tunnel" through the narrow depletion layer, and a large current called Zener current is produced. The process of generating large current using a Zener diode in this way is called Zener breakdown. Zener breakdown is reversible, since Zener breakdown does not destroy the p-n junction.

This means, the Zener current returns to its pre-breaking level when the applied potential difference is reduced below Zener voltage. Zener diodes with specified Zener voltage, V_z are made by controlling the doping. The typical values of Zener voltage are less than 5V or between (5-8) V. The breakdown voltage and the knee point of Zener diode is sharp and well defined as compared to normal diode, Figure 3.12 (b).

3.2. 5 Avalanche breakdown

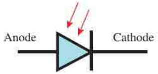
If a reverse voltage larger than the Zener breakdown voltage is applied in a normal diode, the built in electric field at the p-n junction causes the minority charge carriers to move with high speed and colliding with the atoms which knock off more electrons. The produced electrons are again accelerated and collide with other atoms and eventually a large number of free electron is produced in avalanche.

The strong electric field, (E), across the p-n junction will drift the produced electrons across the p-n junction resulting to an abrupt increase of diode current called avalanche current. The sudden increase of current, raises the temperature of the diode, and this may permanently damage the normal diode. However, there is a special type of diode called avalanche diode which is specially designed to operate in the avalanche voltage without damage. The avalanche breakdown occurs at an avalanche voltages greater than 6 V or 8 V.

3.2.6 Photo-diode

A photodiode is a p-n junction diode that absorbs light energy to produce reverse current. Sometimes, a photodiode is also called a photo-detector, a light-detector, or photo-sensor. The circuit symbol for a photodiode is shown in Figure 3.13 (a). Photodiodes are particularly designed to work in reverse bias condition. The reverse current is directly proportional to the intensity of the light falling on it. This means that, the larger the intensity of light the larger the reverse current. Solar cells are considered as large area photodiodes because they convert solar energy into electrical energy.

(a)



(b)

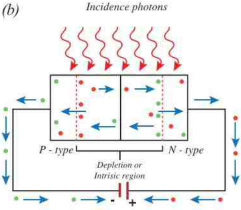


Figure 3.13 Working principle of Photodiode

(a) Working principle of a photodiode

A photodiode is considered as a high-impedance non-ohmic photosensitive device whose current is nearly independent of the applied voltage. Incident light falling on a reverse-biased p-n junction causes bond breakdown. Further increase in light will allow more electrons (and more holes) to flow across the junction, hence a photodiode conducts regardless of its reverse-biased mode as depicted in Figure 3.13 (b). The working principle of a photodiode is based on the photoelectric effect. When a photon of light of desired energy strikes on a photodiode, it is absorbed and an electron-hole pairs is generated. This mechanism of generating electron-hole pairs is called the *internal photoelectric effect*. This effect will generate more free electrons and more holes. The added free electrons will increase the reverse current. As intensity of light incident on the photodiode increases, the reverse current also increases. Therefore, as the incident light intensity increases, the resistance of the photodiode decreases.

(b) Uses of photodiodes

Photodiodes are widely used in electronic circuits such as electronic counters, automatic switching *"on"* and *"off"* of street lights, and in detection of optical signals in the absence and presence of lights. Photodiodes are also used in some devices e.g. cameras, medical devices, safety equipment, optical communication devices, position sensors, barcode scanners, automotive devices, and surveying instruments.

(c) The light-emitting diodes (LED)

A light-emitting diode (LED) is a semiconductor device that emits visible light when an electric current passes through it. The circuit symbol, forward bias as a principle of operation for LED and energy levels of LED are shown in Figure 3.14 (a), 3.14 (b), and 3.14 (c), respectively. The colour of the emitted light depends on the semiconductor material fabricated the diode and the brightness is nearly proportional to the size of forward current. The emitted light colour by LED can range from red (at a wavelength of approximately 700 nm) to blue-violet (about 400 nm). Some LEDs emit infrared (IR) radiation (830 nm or longer); such a device is known as an infrared-emitting diode (IRED). The main semiconductor materials used to manufacture LEDs are indium gallium nitride (InGaN) for blue, green and ultra

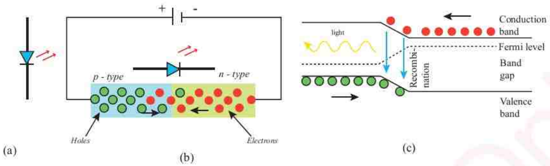


Figure 3.14 Symbol, working principle and energy level of LED

violet LEDs; aluminium gallium indium phosphide (AlGaInP) for yellow and orange LEDs; and aluminium gallium arsenide (AlGaAs), for red and infrared LEDs.

Activity 3.4

Visit any TV, radio and mobile phones workshop and identify devices that make use of any semiconductor components. Draw schematic diagrams of the semiconductor components observed.

(d) Diode as rectifier

Rectification is a process of converting alternating current (*a.c.*) to direct current (*d.c.*). The diode can be used as a rectifier because of the difference in its forward and reverse bias properties. There are two types of rectification, namely *half-wave rectification (HWR)* and *full-wave rectification (FWR)*.

3.2.7 Half wave rectification

Half wave rectification (HWR) can be achieved by using a single diode in which case it conducts current only during the positive half cycle of an *a.c.* During the negative half cycle, no current is conducted, and consequently no voltage across the load.

Mode of operation: During the first half cycle of the sinusoidal waveform as shown in Figure 3.15 (a), at the transformer, terminal A is positive and terminal B is negative. The diode is then forward-biased and current flows around the circuit through the diode, the load resistor R_L back to the transformer winding, as shown in Figure 3.15 (b). During the second half-cycle, terminal A is negative and terminal B is positive, the diode is reverse-biased, and no current flows.

The diode, therefore, conducts on every positive half-cycle of the input and hence, *half-wave rectification* is achieved see Figure 3.15 (c). The output of this rectifier can be improved by putting a large capacitor in parallel with the load. The capacitor is charged during the positive half-cycle of the alternating current and discharges through the load in the negative half-cycle. This action is called smoothing.

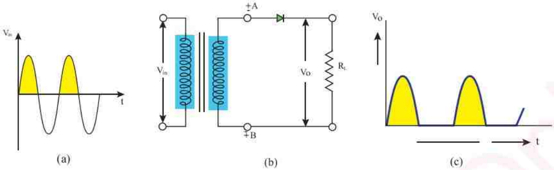


Figure 3.15 A half wave rectification

3.2.8 Full wave rectification (FWR)

In full-wave rectification (FWR) both half cycles of the alternating current are converted to direct current. Two diodes and a centre-tapped transformer as shown in Figure 3.16, or four diodes in a bridge configuration as illustrated in Figure 3.17, can be used.

Mode of operation: In the case of the circuit for full wave rectification Figure 3.16(b), the voltage at point A does the opposite of the voltage at point B, whereas, voltage at point A increases in a positive direction, voltage at point B increases in a negative direction, Figure 3.16 (a). In the positive half-cycle, point A is positive with respect to O and diode D_1 is forward biased, whereas diode D_2 is reverse-biased. Therefore current flows through D_1, R_L, and back to O. In the negative half-cycle, point B is positive with respect to O. Diode D_2 is forward biased, whereas diode D_1 is reverse-biased. In this case, current flows through D_2, R_L and back to O. The direction of the current through R_L is the same as in the first half-cycle, Figure 3.16 (c).

Full wave rectification can also be achieved by using four diodes. The circuit and the resulting full wave rectification using a bridge rectifier is shown in Figure 3.17 (a) and 3.17 (b) respectively. It can be noted

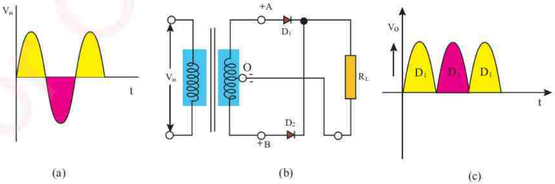


Figure 3.16 Full wave rectification, using two diodes

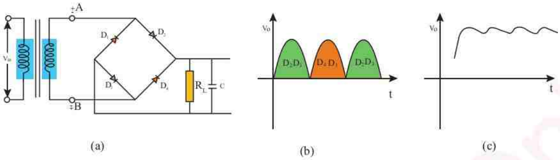


Figure 3.17 Full wave rectification using four diodes

from the Figure 3.17 that, during the first half cycle, point A is positive and point B is negative, so that diodes D_2 and D_3 are forward-biased and D_1 and D_4 are reverse-biased. Diodes D_2 and D_3 conduct and current flows from A via diode D_2, resistor R_L, diode D_3 and back to the source at point B. In the second half-cycle, B is positive and point A is negative, diodes D_1 and D_4 are forward-biased while D_2 and D_3 are reverse- biased. Diodes D_1 and D_4 conduct, and the current flows from B via diode D_4, R_L, diode D_1 and back to the source at point A.

The current through R_L is in the same direction in both half-cycles. The output wave in the bridge rectifier is the same as that for the centre tapped transformer with two diodes full-wave rectifier. When a capacitor is connected across the load R_L, very little ripples in the output *d.c.* signal are obtained as depicted in Figure 3.17 (c). The capacitor C, is said to smoothen the output *d.c.* signals.

3.2.9 Zener diode as a voltage stabilizer

The Zener diode can be used as a *d.c.* voltage regulator to provide a constant voltage output. The circuit diagram for this arrangement is shown in Figure 3.18, where the resistor R_S is connected in series with the Zener diode and the voltage source, V_{in} being connected across the combination. The negative terminal of a Zener diode is connected to the positive terminal of the *d.c.* supply, so it is reverse biased and will be operating in its breakdown condition. The stabilized output voltage, V_{out} is taken across the Zener diode. The resistor, R_S is chosen to limit the maximum current flowing in the circuit.

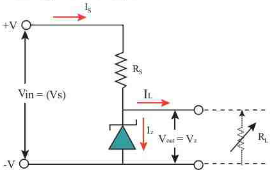


Figure 3.18 Zener diode voltage stabilizer circuit

In the absence of the load current will be zero, that is, and all the circuit current flows through the Zener diode, which in turn, dissipates its maximum power. If a value of the series resistor is small, this will result in a larger diode current when a large load resistance, is connected. This large value of diode current will increase the power dissipation requirement of the diode. Care must be taken when selecting the appropriate value of series resistance so that the Zener's maximum power rating is not exceeded under no-load or high-impedance load condition.

The load resistor is connected in parallel with the Zener diode, then the voltage across R_L (V_R) is always the same as the Zener voltage (V_Z). There is a minimum Zener current for which the stabilization of the voltage is effective. The Zener current must stay above this value when operating under load within its breakdown region at all times. The upper limit of the current depends upon the power rating of the device. The supply voltage, V_S, must be greater than V_Z.

Limitation of Zener diode stabilizer circuits is that, the diode can sometimes generate electrical noise on top of the *d.c.* supply as it tries to stabilize the voltage. Normally, this is not a problem for most applications; nevertheless, the addition of a large value decoupling capacitor across the Zener's output may be required to give additional smoothing to the output voltage.

Activity 3.5

Given the following devices; *a.c.* power supply, transformer, diodes, connecting wires, display unit (oscilloscope), resistors, and capacitors, design and construct circuits for half and full wave rectifiers.

Example 3.1

A stabilized power supply of $6.0\,V$ is required to be produced from 15 V, *d.c.* power supply input source. The maximum power rating P_Z of the Zener diode is 3 W. Use the Zener diode regulator circuit in Figure 3.18 to calculate:

(a) The maximum current flowing through the Zener diode.

(b) The minimum value of the series resistor, R_S

(c) The load current, I_L if a load of 2 kΩ is connected across the Zener diode.

(d) The Zener current, I_Z at full load.

Solution

(a) The maximum current flowing through the Zener diode:

$$I_S = I_{\max} = \frac{P_{\max}}{V} = \frac{3\,\text{W}}{6\,\text{V}} = 0.5\,\text{A}$$

(b) The minimum value of the series resistor, R_S:

$$R_s = \frac{V_s - V_z}{I_z} = \frac{15\,\text{V} - 6\,\text{V}}{0.5\,\text{A}} = 18\,\Omega$$

(c) The load current I_L if a load resistor of 2 kΩ is connected across the Zener diode

$$I_L = \frac{V_z}{R_L} = \frac{6\,\text{V}}{2\times10^3\,\Omega}$$

$$= 0.003\,\text{A} = 3\,\text{m A}$$

(d) The Zener current I_Z at full load.

$$I_Z = I_S - I_L = (0.5\text{-}0.003)\,\text{A}$$

$$= 497\,\text{m A}$$

Exercise 3.2

1. What are semiconductor materials? Distinguish between intrinsic and extrinsic semiconductors.
2. Describe the formation of n-type and p-type semiconductors.
3. What is a p-n junction? Briefly describe the formation of potential barrier of a p-n junction diode.
4. With the aid of diagrams, explain the forward and reverse biased mode of p-n a junction. Why is a p-n junction called a semiconductor diode?
5. Draw and explain the I - V curve characteristic of a p-n junction.
6. Why is silicon preferred to germanium in the manufacturing of semiconductor devices?
7. Distinguish between a light emitting diode (LED) and a photodiode.
8. The circuit in Figure 3.19 shows two diodes each with a forward resistance of 50 Ω and with infinite reverse resistance. If the battery voltage is 6 V, what is the current through the 100 Ω resistor?

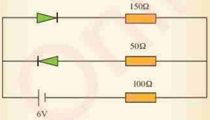


Figure 3.19: Resistors – diodes circuit

9. In Figure 3.18, what is the voltage needed to maintain 15 V across the load resistance R_L of 5.0 kΩ? Assuming the series resistor R_s is 400 Ω and the Zener diode requires a minimum current of 10 mA to work satisfactorily, what is the required Zener diode rating?

3.3 Transistors

When a third layer, either n-type or p-type is added to a semiconductor diode and forming two junctions, the resulting device is known as transistor. A transistor is a three layer semiconductor device consisting of either two n-type and one p-type layers or two p-type and one n-type layers of material. The type of transistor when n-type layer fused to p-n diode is called npn transistor, while p-type layer fused is called pnp transistor as illustrated in Figures 3.20 (a) and 3.20 (b) with its symbols respectively. An example of a physical structure of a transistor is shown in Figure 3.20 (c).

The word transistor is a prefix for 'trans' which means the signal transfer property of the device, and 'istor' classifies it as a solid state element in the same general family as resistors. The npn and pnp transistors are called *bipolar junction transistors* (BJT) because their conductions involve two charge carriers, holes and free electrons. Transistors can also be constructed using either silicon or germanium, but virtually all transistors other than exotic types use silicon; the exotic types use compound semiconductors such as gallium arsenide. An arrow in the transistor symbol indicating conventional flow of current.

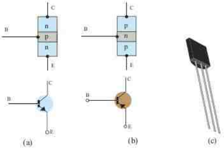
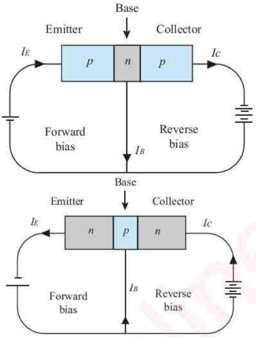


Figure 3.20 npn and pnp transistors

3.3.1 Naming of transistor terminals

The transistor has three terminals, namely Emitter (E), Base (B), and Collector (C). In practice, these three terminals are grown on the same crystal by doping process (additional of corresponding layer). The emitter is heavily doped, so that it emits charge carriers to the collector. The base is thin and lightly doped in order to allow charge carriers from the emitter to pass through it without recombination. The collector is wider and moderately doped, in order to collect most of the charge carriers from the emitter. The transistor consists of two junctions namely, the emitter-base junction and the collector-base junction. For proper operation of the bipolar junctions transistor, the emitter-base junction is always forward biased and the collector-base junction is always reverse biased as illustrated in Figure 3.21.



(b)

Figure 3.21 The pnp and npn transistor biasing

3.3.2 Mode of operation of transistor

Transistors are used in analog circuits to amplify a signal. They are also used in power supplies as regulators and as switches in circuits. Figure 3.22 (a) shows npn transistor connected to two power supplies, one to the emitter and the other to the collector. Since the emitter-base junction is forward biased, the free electrons in the emitter will be repelled towards the collector through the base; this will establish an emitter current (I_E).

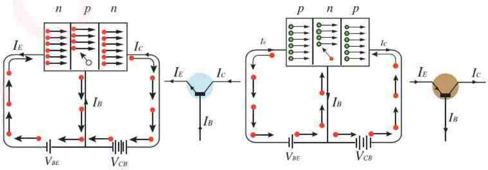


Figure 3.22 The working principle of npn and pnp transistors

In the base, a few electrons recombine with holes in it and majority of the free electrons will pass to the collector. The collector-base junction is reverse biased, the free electrons in the collector will be attracted by the battery to constitute a collector current (I_C). In the base, few electrons diffuse into the base leading to formation of a base current (I_B). The operation of the *pnp* transistor is exactly the same, if the roles played by the electrons and holes are interchanged as illustrated in Figure 3.22 (b). By Kirchhoff's current law the following formula holds true for transistors:

$$I_E = I_B + I_C \qquad (3.1)$$

It is noted that, the potential difference across the base-emitter, V_{BE} is always less than the potential difference across the base-collector, V_{BC} for the current to flow. The importance of transistor action is that the input circuit, (emitter-base junction) has a low resistance because of forward bias while the output circuit (collector-base junction) has a high resistance because of reverse bias. Therefore, a transistor transfers the input current signal from a low resistance to high resistance circuit. This is the key factor responsible for the amplifying action of a transistor. Note that potential drop and the current directions for npn and pnp transistors are shown in Figure 3.23.

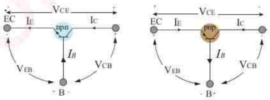


Figure 3.23 The potential drop and current directions

Activity 3.6

(i) Open up any electronic device such as a computer, radio, TV and phone. Identify any bipolar junction transistor(s) present.

(ii) Use a multimeter to identify base, collector and emitter terminals of transistors.

(iii) Use a multimeter to differentiate pnp from npn transistors.

3.3.3 Transistor circuit configurations

In the previous subsection it was noted that a transistor has three terminals but when it is connected in a circuit there is a need of four terminals, two for output and two for input. The fourth terminal is obtained by making one terminal of the transistor common to input and output terminals. The input is applied between this common terminal and one of the other two terminals while the output is found between the common terminal and remaining terminals. The transistor can be connected in three different configurations namely, *common base* (CB), *common collector* (CC) and *common emitter* (CE). The namings of these configurations depends on which terminal of the transistor is grounded. Hence it can be either base, collector or emitter is grounded.

(a) Common base (CB) configuration

This transistor configuration provides a low input impedance while offering a high output impedance. Although the voltage output of CB is high, the current gain is low and the overall power gain is also low when compared to the other transistor configurations. In this configuration, the input and output signals are always in phase as illustrated in Figure 3.24.

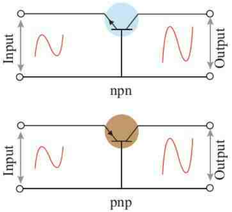


Figure 3.24 The common base circuit configuration

For the common base amplifier circuit, in both npn and pnp circuits, the input is applied to the emitter, and the output is taken from the collector, and the common terminal for both circuits is the base.

The common base amplifier configuration is not used widely as transistor amplifier. However, it finds uses as an amplifier that requires low input impedance levels. It is commonly used in moving-coil microphone, pre-amplifiers since these devices requires very low impedance levels.

(b) The voltage gain and power gain in the CB configuration

CB configuration circuit for voltage gain (A_v) determination is shown in Figure 3.25. This type of configuration has a high ratio of output to input resistance that is load resistance R_L to input resistance R_{in}.

For npn transistor, it has been noted that electrons are emitted from n-emitter towards the p – base, a constant fraction of electrons, α, will reach the n – collector (typically $\alpha \approx 0.98$). For a common base configuration the voltage gain (A_V) is given by:

$$A_v = \frac{V_{out}}{V_{in}} = \frac{I_C \times R_L}{I_E \times R_{in}} \qquad (3.2)$$

where $\frac{I_C}{I_E} = \alpha$ (current gain)

$$A_v = \alpha \frac{R_L}{R_{in}} \qquad (3.3)$$

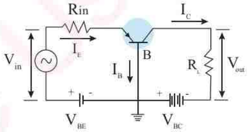


Figure 3.25 Voltage gain of a common base configuration

The power gain is defined as a product of voltage gain and the current gain.

$$Power\ gain = \alpha \times A_v = \alpha^2 \times \frac{R_L}{R_{in}} \qquad (3.4)$$

Generally the common base circuit is used only in single stage amplifier circuits, such as microphone pre-amplifiers or radio frequency (R_f) amplifiers due to its good response to high frequency. It is worth noting that the current gain of a common-base amplifier is always less than one. However, the voltage gain is a function of input and output resistances and also the internal resistance of the emitter-base junction.

(c) Common emitter (CE) configuration

In a common emitter configuration, the emitter is grounded. The input signal is applied between the base and the emitter, while the output is taped between the collector and the emitter as shown in Figure 3.26. The common emitter configuration is the most widely used due to it's best combination of current gain and voltage gain. The circuit provides low input and high output impedance levels leading to both larger current and voltage gains. It should be noted that the output signal is out of phase with input signal, as illustrated in Figure 3.26 for both npn and pnp transistors .

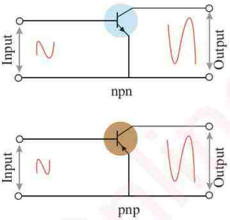


Figure 3.26 The common emitter circuit configuration

The disadvantage of this circuit is that the leakage current is amplified in the circuit. This challenge can be removed by bias stabilazation methods.

(d) Current gain and power gain in the CE configuration

The common emitter amplifier configuration produces the highest current among all the three configurations. This is because the input impedance is low as it is connected to a forward biased p-n junction, while the output impedance is high as it is taken from a reverse biased p - n junction.

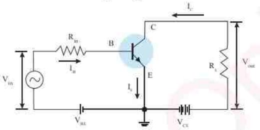


Figure 3.27 The common emitter transistor circuit

In this type of configuration, the current flowing out of the transistor must be equal to the currents flowing into it. By Kirchhoff's current law:

$$I_E = I_B + I_C$$

If the load resistance (R_L) is connected in series with the collector, the current gain (β) of the common emitter transistor configuration is quite large because I_C is much larger than I_B. The transistor current gain, (β), is given as:

$$\beta = \frac{I_C}{I_B} \qquad (3.5)$$

Small changes in the current flowing in the base will thus control the current in the emitter collector circuit. Typically, β has a value between 20 and 200 for most general purpose transistors. If a transistor has a β value of say 100, then one electron will flow from the base terminal for every 100 electrons flowing between the emitter collector terminals.

Combining equations for both, alpha (α), and beta (β), the mathematical relationship between α, and β are given as:

from $\beta = \frac{I_C}{I_B} \Rightarrow I_C = \beta I_B$ and

$\alpha = \frac{I_C}{I_E} \Rightarrow I_C = \alpha I_E$, but $I_E = I_B + I_C$

Then, the relationship between α and β are either

$$\alpha = \frac{\beta}{\beta + 1} \quad (3.6)$$

or $$\beta = \frac{\alpha}{1 - \alpha} \quad (3.7)$$

In CE configuration voltage gain (A_V) is given by:

$$A_v = \frac{V_{out}}{V_{in}} = \frac{I_C R_L}{I_B R_{in}} = \beta \frac{R_L}{R_{in}}$$

Power gain in C-E configuration is given as a product of current gain and voltage gain.

Thus, power gain $= \beta \times A_v \quad (3.8)$

$$= \beta \times \beta \frac{R_L}{R_{in}} = \beta^2 \frac{R_L}{R_{in}}$$

(e) Common collector (CC) configuration

In common collector configuration (collector is grounded) the collector is common across the supply, and the input signal is connected directly to the base, while the output is taken from the emitter load as shown in Figure 3.28. This transistor configuration is also known as the *emitter follower* because the emitter voltage follows that of the base. The voltage gain of a common collector configuration is unity while its current gain is high. The output and input impedances are low and high respectively. The input and output signals are in phase as shown in Figure 3.28. The emitter follower circuit provides an ideal buffer stage, and as a result it is used in many circuits like an oscillator, where it is not necessary to load a circuit. It also provides a low output impedance to amplifier stages. Due to low output impedance, the common collector configuration is used as a final stage in the designing of an operational amplifier with low output impedance.

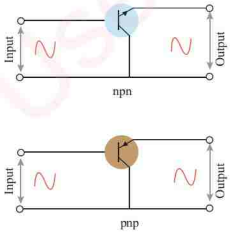


Figure 3.28 The common collector circuit configuration

(f) The current gain and power gain in CC configuration

In the common collector configuration, the load resistance is connected in series with emitter. So its current is equal to that of the emitter current which is given by: $I_{R_L} = I_E = I_B + I_C$ as shown in Figure 3.29.

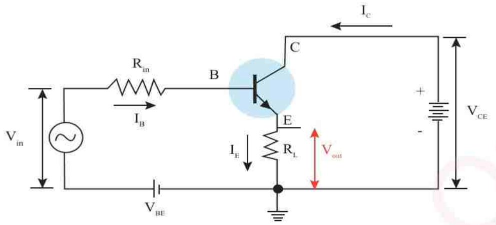


Figure 3.29 The common collector transistor circuit

The emitter current is the combination of the collector and the base current. The load resistance in this configuration has both the collector current and the input current of the base flowing through it. Then the current gain γ of the circuit is given as;

$$\text{Current gain } (\gamma) = \frac{I_E}{I_B} \tag{3.9}$$

$$\gamma = \frac{I_E}{I_B} = \frac{I_B + I_C}{I_B} = 1 + \frac{I_C}{I_B}$$

$$\gamma = 1 + \beta \tag{3.10}$$

The load resistance of the common collector transistor receives both the base and collector currents giving a larger current gain (as with the common emitter configuration), and therefore, providing good current amplification with very little voltage gain.

Voltage gain is given by:

$$A_v = \frac{V_{out}}{V_{in}} = \frac{I_E R_L}{I_B R_{in}} = \gamma \frac{R_L}{R_{in}}$$

$$= (1+\beta)\frac{R_L}{R_{in}} \approx \beta \frac{R_L}{R_{in}};\ 1 \ll \beta$$

The power gain in the CC configuration is given by:

$$\text{Power gain} = \gamma \times A_v$$

$$= (1+\beta) \times A_v \tag{3.11}$$

but for a large β then $1+\beta \approx \beta$

Therefore, power gain in the CC

configuration is approximately equal to

$$\beta A_V = \beta^2 \frac{R_L}{R_{in}}$$

3.3.4 Transistor characteristics

To study the characteristics of the transistor input, output and transfer characteristics a voltage or current applied to one pair of the transistor's terminals is varied and the current or voltage across another pair of terminals can be monitored and recorded. Figure 3.30 depicts a circuit diagram for studying the transistor characteristics of a common emitter configuration (CE).

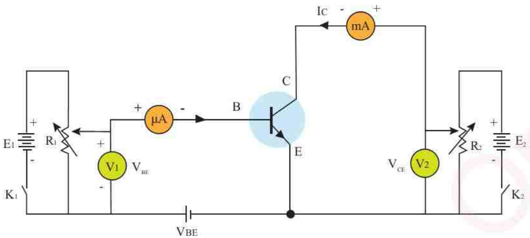


Figure 3.30 CE circuit for investigating the transistor characteristics

(a) Input characteristics

From the circuit in Figure 3.30 the following is implemented: Keeping the collector emitter (V_{CE}) voltage constant, the base emitter (V_{BE}) voltage is increased from zero and the corresponding base current (I_B) values are recorded. This is repeated for increasing values of V_{CE}. The family of curves obtained by plotting I_B against V_{BE} for each value of V_{CE} as shown in Figure 3.31 is called *input characteristics*. The shape of the curves is like that of a forward biased junction diode. I_B is very small for values of the V_{BE} which are less than the contact potential (about 0.6 V for silicon). The input resistance, r_i is defined by:

$$r_i = \frac{\Delta V_{BE}}{\Delta I_B} \text{ at constant } V_{CE} \qquad (3.12)$$

$r_i > 10^3\,\Omega$

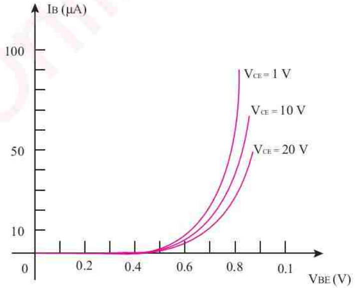


Figure 3.31 Input characteristics of a common emitter configuration

(b) Output characteristics

Referring to circuit in Figure 3.30, keeping the base current (I_B) constant, collector emitter voltage (V_{CE}) is varied and the corresponding collector current (I_C) values are recorded. This is repeated for increasing values of I_B. The family of curves obtained by plotting I_C against V_{CE} for each value of I_B as shown in Figure 3.32 is called *output characteristics*.

For values of V_{CE} which are greater than about 1V, I_C increases only slightly with V_{CE} but it is strongly dependent on I_B. Transistors that are used as amplifiers are operated at values of V_{CE} which are to the right of the knee volts (about 1 V). The curves in this region are linear and an amplifier produces undistorted output.

The output resistance at constant I_B is given by:

$$r_0 = \frac{\Delta V_{CE}}{\Delta I_C} \qquad (3.13)$$

$$r_0 > 10^5\Omega$$

The value of the output resistance depends on the particular value of I_B, it also depends on V_{CE} because I_C varies linearly with V_{CE} at this region just to the right of the knee volt.

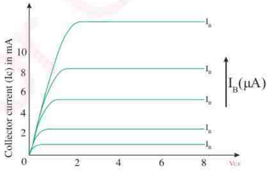


Figure 3.32 Output characteristics of a common emitter configuration

(c) Transfer characteristics

From a circuit depicted by Figure 3.30, if a value of V_{CE} is fixed and input resistance, R_1 is varied then the different values of a collector current (I_C) and base current (I_B) are obtained. A plot of I_C against I_B gives a linear graph showing that I_C is directly proportional to $I_{B,}$ as illustrated in Figure 3.33. The characteristic also shows that when I_B is zero, I_C is a small value about $0.01\propto A$ for silicon and 2μA for germanium at 15 ϒC. This is called the *leakage current* $\left(I_{CE0}\right)$ and is due to minority charge carriers crossing the reverse bias collector –base junction to emitter through collector and base. Therefore the current gain is given as:

$$\beta = h_{fe} = \frac{\Delta I_C}{\Delta I_B} \qquad (3.14)$$

where h_{fe} is small signal forward current transfer ratio while β is current gain.

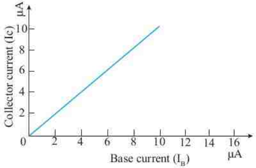


Figure 3.33 Transfer characteristics of a transistor

(d) The effect of temperature on transistor performance

Semiconductors are very sensitive to temperature variations. When a transistor is in operation, it produces heat, and almost the entire heat is produced at the collector-base junction. The heating dissipates power from the transistor and causes temperature of the junction to rise. When the temperature of a transistor rises, the thermal energy generates more free electrons and holes that add to the collector current. The increased collector current will cause further heating of the transistor, which may lead to transistor damage. This process is known as *thermal runway* and is a cause of transistor damage in devices such as power amplifiers. Fortunately it is relatively easy to stabilize the situation by using techniques such as negative feedback that will be discussed in section 3.5. Also the initial temperature rise of transistors can be associated with an increase in the surrounding temperature, and a need of replacing the transistor with another which has a larger current gain (operating point shift) is necessary.

3.3.5 Design and construction of transistor circuits

This subsection will describe the design and construction of transistor amplifiers and switches. Practical transistor amplifiers operate using *a.c.* signal inputs which alternate between a positive and negative values. The amplifier circuit should be designed so as to operate between the two maximum levels of the input signal. This is achieved by using a process known as *biasing*. Biasing is very important as it establishes the correct operating point of the transistor amplifier, and if correctly done it will avoid distortion of the output signal.

(a) Single stage (CE) amplifier

The most common amplifier configuration for npn transistor is that of the common emitter amplifier circuit. In designing an amplifier, the first step is to provide necessary base bias voltage to the transistor. This can be achieved by self-bias or potential divider bias methods. The potential divider transistor bias is shown in Figure 3.34. This method of biasing the transistor, greatly reduces the effects of varying current gain (β) by holding the base bias at a constant steady voltage level allowing for best stability. The quiescent base voltage (V_B) is determined by the potential divider network formed by the two resistors, R_1 and R_2, and the power supply voltage V_{CC}.

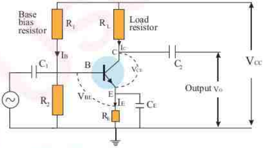


Figure 3.34 The npn transistor in a common emitter configuration circuit

Different components and their functions in the CE amplifier circuit, as in Figure 3.34 are as outlined below:

(i) A coupling capacitor C_1: blocks *d.c.* components in the input signal entering the circuit while a coupling capacitor C_2 blocks any *d.c.* components in the output signal. This ensures that the bias condition set up for the circuit to operate correctly it will not affect any additional amplifier stages. The capacitors will pass *a.c.* signals only and block any *d.c.* components hence maintaining a good operating point stability.

(ii) A bypass capacitor C_E blocks undesirable feedback of the amplified signal to the base emitter circuit. However, this bypass capacitor short-circuits the emitter resistor, R_E at high frequency signals. The bypass capacitor C_E is chosen to provide a reactance of at most $\frac{1^{th}}{10}$ of the value of R_E at the lowest operating signal frequency.

(iii) R_1 and R_2 forms a potential divider which provides the necessary base bias voltage to the circuit.

(iv) Load resistance, R_L produces the output of the transistor.

(v) Emitter resistance, R_E is used to stabilize the circuit in case of excessive temperature rise.

The voltage level generated at the junction of resistors R_1 and R_2 holds the base voltage (V_B) constant at a value below the supply voltage. Then the potential divider network used in the common emitter amplifier circuit divides the input signal in proportion to the resistance. This bias reference voltage can be easily calculated using the simple voltage divider formula as:

$$V_B = \frac{R_2}{R_1 + R_2} V_{CC} \qquad (3.15)$$

From Figure 3.34 in the absence of *a.c.* signal, the C_1 capacitor provides very high impedance (open circuit). In this case the *d.c.* equivalent circuit for common emitter amplifier becomes as shown in the Figure 3.35.

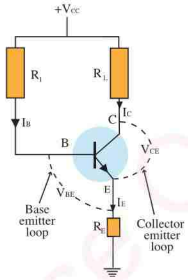


Figure 3.35 An equivalent circuit of a common emitter amplifier

Using Kirchhoff's voltage law to the base emitter loop, the input voltage is given as:

$$V_{CC} = I_B R_1 + V_{BE} + I_E R_E \qquad (3.16)$$

Applying Kirchhoff's voltage law to the collector emitter loop, the output equation is given as:

$$V_{CC} = I_C R_L + V_{CE} + I_E R_E \qquad (3.17)$$

(b) Quiescent state (operating point)

The common emitter circuit shown in Figure 3.36 can be used to amplify small voltage changes. A small voltage source provides both the base emitter bias and the collector-emitter bias a feature which is made possible by the presence of the base bias resistor. Denote collector current as I_C and voltage drop across collector emitter as V_{CE}. If I_C is the collector current when no signal is applied, the value of V_{CE} is such that the output signal can swing by equal amount above and below this value without driving the transistor into **saturation region** (V_{CE} = 0 and I_C = maximum) or **cut off region** (I_C = 0 and V_{CE} = maximum, i.e. V_{CE} is equal to V_{CC}).

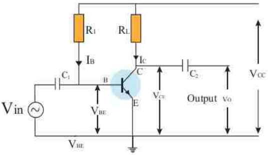


Figure 3.36 A simple amplifier circuit

Using Kirchhoff's voltage law, the output voltage is given by;

$$V_{cc} = I_C R_L + V_{CE} \qquad (3.18)$$

At saturation point, V_{CE} = 0, then the saturation collector current, I_C is maximum, given by:

$$I_C = \frac{V_{cc}}{R_L} \qquad (3.19)$$

The component values are chosen so that the quiescent collector emitter voltage V_{CE} is about half the battery voltage V_{CC}, i.e.

$$V_{CE} = \frac{1}{2}V_{cc} \qquad (3.20)$$

Once the supply voltage V_{CE} and collector current I_C at operating point are decided, collector resistance R_L can be calculated as:

$$R_L = \frac{V_{CE}}{I_C} \qquad (3.21)$$

Whenever a component value R_L has been calculated, it is unlikely that the result of the calculation will match any of the available preferred values of standard resistors. Therefore, it will be needed to choose the nearest preferred value of standard resistors. To provide efficient bias stabilization, the collector emitter voltage V_{CE} should be about 10% to 15% of V_{CC}. Assuming that I_E is the same as I_C (they differ by a small amount of the base current), then the value of the emitter resistor R_E can be calculated as;

$$R_E \approx \frac{V_{CE}}{I_E} \approx \frac{V_{CE}}{I_C} \qquad (3.22)$$

To estimate the value of the base current I_B, the collector current I_C is divided by the given transistor current gain β obtained from the data sheet for the particular transistor. Because the β varies from one transistor to another, it may be quoted as a typical value or as a range between minimum and maximum values.

The value of β also varies with collector current. Whatever value you choose for β, the result of calculated I_B will be an approximation, so the base voltage will probably not be accurate. However, this can be "fine-tuned" when the amplifier is being constructed. The base voltage should be about 0.6 V higher than V_{CE} for silicon transistor to ensure that the input signal is biased on the linear part of the transistor input characteristic curves.

Typical junction voltages for silicon and germanium transistors are given in Table 3.1.

Table 3.1: Typical junction voltages for silicon and germanium transistors

Transistors	Saturation region		Active region		Cut-inn / off region	
	V_{CE}	V_{BE}	V_{BE}	V_{CE}	V_{BE}(cut-inn)	V_{BE}(cut-off)
Si	0.2 V	0.8 V	0.7 V	0.3 V	0.5 V	0 V
Ge	0.1 V	0.3 V	0.2 V	0.2 V	0.1 V	-0.1 V

(c) Load line and the *d.c.* operating point

A load line is a line whose voltage–current plot represents a load resistance, and it is drawn over the output characteristics (I_C–V_{CE} curve) as shown in Figure 3.37.

The load line cuts any of the I_B lines and a value of V_{CE} can be read off, so that the output voltage swing for a given input current swing can be evaluated. The *d.c.* load line can be drawn to show all the possible operating points of the transistor from fully "ON" to fully "OFF", and to which the quiescent operating point or **Q-point** of the transistor amplifier can be found.

From the equation (3.18) it follows that:

$$I_C = -\frac{V_{CE}}{R_L} + \frac{V_{CC}}{R_L}$$

At saturation point, the collector current is maximum for a given base current and thus $V_{CE} = 0$, then:

$$I_c = \frac{V_{CC}}{R_L}$$

At cutoff point, the collector current is negligible as it is essentially due to minority carriers and that $V_{CE} = V_{CC}$.

The graph of I_C against V_{CE} is a straight line with a slope $-\frac{1}{R_L}$

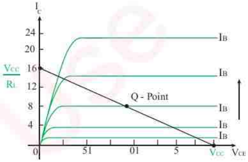


Figure 3. 37 A load line and Q-point

The aim of any small signal amplifier is to amplify all of the input signals with a minimum amount of distortion to the output signal. The operating quiescent point needs to be correctly selected in order to obtain small distortion when a transistor is used as an amplifier.

This is in fact the *d.c.* operating point of the amplifier, and its position may be established at any point along the load line by a suitable biasing arrangement. The best possible position for this Q-point should be close to the centre position of the load line as reasonably as possible, for its swing peak to peak, in order to ensure proper operation of a transistor as an amplifier.

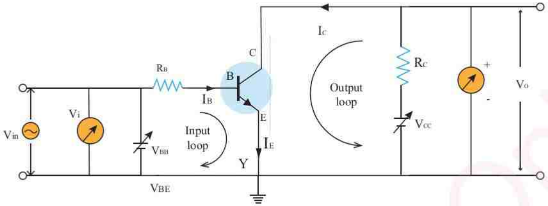


Figure 3.38 A circuit of a transistor as a switch

3.3.6 Transistor as a switch

A transistor acts as a switch when operated at cut-off and saturation points. At cut-off point, a transistor acts as an open switch and thus it becomes OFF switch, and at saturation point, it acts as a closed switch and thus it becomes ON switch. To understand the operation of the transistor as a switch the analysis of the behaviour of the base-biased transistor in CE configuration is considered as shown Figure 3.38.

Applying Kirchhoff's voltage law to the input and output loops of this circuit, we get

Input: $V_{BB} = I_B R_B + V_{BE}$ 3.23

Output: $V_{CE} = V_{CC} - I_C R_C$

Let V_{BB} be the *d.c.* input voltage V_i and V_{CE} be the *d.c.* output voltage V_o. The above equations become:

$$V_i = I_B R_B + V_{BE} \text{ and } V_0 = V_{CC} - I_C R_C$$

The variation of V_o as V_i increases

***Case I*:** For silicon (Si) transistor, as long as input V_i is less than 0.6 V, the current I_C will be zero; therefore:

$$V_o = V_{CC}$$

In this case the transistor will be in cut off state (region).

***Case II*:** When V_i becomes greater than 0.6 V the silicon transistor is in active state with some current I_C in the output path and the output V_0 decrease as $I_C R_C$ increases. With increase of V_i, I_C increases almost linearly Figure 3.39 (a) and so V_o decreases linearly to about 1.0 V. Further increase in V_i implies V_o decrease towards zero though it may never become zero; the variation becomes non-linear and transistor is said to be in saturation state.

A plot of V_o against V_i, which is called the transfer characteristics of the base-biased transistor, as shown in Figure 3.39 (b), depicts the Silicon transistor behaviour between saturation and cut off states. Slope of the linear portion in the active region gives rise to the voltage gain, $A_v = \frac{\Delta V_o}{\Delta V_i}$

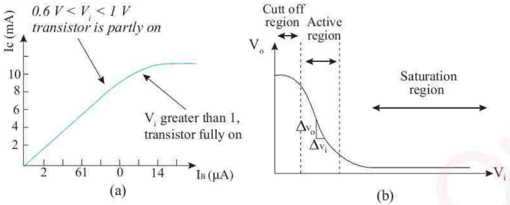


Figure 3.39 Voltage and current transfer characteristics of transistor as a switch

In summary, as long as V_i is *low* and unable to forward-bias the transistor, V_o is *high*. If V_i is *high* enough to drive the transistor into saturation, then V_o is *low*, near to zero. The transistor is said to be *switched off.* It is not conducting, and it is said to be *switched on* when it is driven into saturation. The *low* V_i to the transistor gives a *high* V_o and a *high* V_i gives a *low* V_o. When the output $V_o = V_{CC}$ then all the battery voltage appears across the transistor which behaves as a very large resistor and is said to be at cut-off state (region). Switching circuits are designed in such a way that the transistor does not remain in active state (region).

Activity 3.7

(i) Design a transistor circuit using a light dependent resistor to switch on a light when it gets dark and vice versa.

(ii) Design a transistor circuit using a temperature sensor e.g. a thermistor, to switch on a fan when it gets hot and vice versa.

Example 3.2

For a CE transistor amplifier, the audio signal voltage across the collector resistance of 2.0 k Ω is 2.0 V. Suppose the current amplification factor of the transistor is 100. What should be the value of R_B in series with V_{BB} supply of 2.0 V, if the *d.c.* base current has to be 10 times the base signal current?

Solution

The output *a.c.* voltage is 2.0 V. The *a.c.* collector current becomes:

$$I_c = \frac{V}{R_L} = \frac{2\text{V}}{2000\ \Omega} = 1\ \text{mA}$$

The signal current through the base is given by:

$$I_B = \frac{I_C}{\beta} = \frac{1\ \text{mA}}{100} = 0.01\ \text{mA}$$

The *d.c.* base current has to be

$$I_{dc} = 10 \times 0.01 = 0.1\ \text{mA}$$

Using equation 3.23

$$R_B = \left(\frac{V_{BB} - V_{BE}}{I_B}\right) = \left(\frac{2\text{V} - 0.6\text{V}}{0.1}\right) = 14\ \text{k}\Omega$$

Example 3.3

In Figure 3.38, the V_{BB} supply can be varied from 0 V to 6.0 V. The silicon transistor has $\beta = 100$ and $R_B = 10$ kΩ, $R_C = 4$ kΩ, $V_{CC} = 6.0$V. Assume that when the transistor is saturated, $V_{CE} = 0$ V and $V_{BE} = 0.7$ V.

(i) Calculate the minimum base current, for which the transistor will reach saturation.

(ii) Determine V_i when the transistor is switched on.

(iii) Find the range of V_i for which the transistor is switched off and switched on.

Solution

(i) At saturation point, from equation (3.19),

$$I_C = \frac{V_{cc}}{R_C} = \frac{6}{4\times 10^3} = 1.5 \text{ mA}$$

From equation (3.5) the base current is

$$I_B = \frac{I_c}{\beta} = \frac{1.5\text{mA}}{100} = 15 \propto\text{A}$$

(ii) The input voltage at which the transistor will go into saturation is given by:

$$V_{BB} = I_B R_B + V_{BE}$$

$$= 15\text{mA}\times 10 \text{ k}\Omega + 0.7 \text{ V} = 0.85 \text{ V}$$

(iii) The range of the value of input voltage below which the transistor remains cut-off is given between $V_{iL} = 0.6$ V and $V_{iH} = 0.85$ V, the transistor remains in active state, between 0.0 V and 0.6 V, the transistor will be in switched off state. Between 0.85 V and 6.0 V, it will be in switched on state.

Example 3.4

In the circuit shown in Figure 3.40.

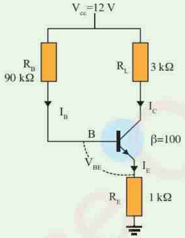


Figure 3.40 Transistor circuit

(i) Show that a silicon transistor will operate in the saturated region.

(ii) Calculate the value of R_E so that the transistor is just out of saturation i.e. in active region.

Solution

(i) At saturation

Applying KVL for the input loop (base emitter loop)

$$V_{CC} = I_B R_B + V_{BEsat} + I_E R_E$$

Similarly to the output loop (collector emitter loop)

$$V_{CC} = I_c R_L + V_{CEsat} + I_E R_E$$

Remember:

$$I_E = I_B + I_C$$

$$V_{CC} = I_B\left(R_B + R_E\right) + V_{BEsat} + I_C R_E$$

$$12 = I_B\left(90+1\right)\times 10^3 + 0.8 + 10^3 I_C$$

$$91\times 10^3 I_B + 10^3 I_C = 11.2 \qquad (i)$$

Similarly

$$V_{CC} = I_c(R_L + R_E) + V_{CEsat} + I_B R_E$$

$$12 = I_C(3+1)\times 10^3 + 0.2 + 10^3 I_C$$

$$4\times 10^3 I_C + 10^3 I_B = 11.8 \qquad (ii)$$

Solving simultaneous equations (i) and (ii) then

$$I_C = 2.93\text{ mA}$$

$$I_B = 91\,\mu\text{A}$$

To justify the transistor is saturated, evaluate:

$$I_B > \frac{I_C}{\beta} \Rightarrow I_B > \frac{2.93\times 10^{-3}}{100} = 2.93\times 10^{-5}\text{A}$$

$$= 0.239\ \propto\text{A}$$

Hence $\frac{I_C}{\beta}$ is less than I_B and this justifies the assumption that the transistor is saturated.

(ii) *Out of saturation (active region)*

Use $I_C = \beta I_B$ *or* $I_E = (1+\beta)I_B$

Applying KVL on base section,

$$V_{CC} = I_B R_B + V_{BEact} + I_E R_E$$

Similarly at the output side,

$$V_{CC} = I_c R_L + V_{CEact} + I_E R_E$$

$$V_{CC} = I_B R_B + V_{BEact} + (1+100) I_E R_E$$

$$12 = 90\times 10^3 I_B + 0.7 + (1+100) I_B R_E$$

$$90\times 10^3 I_B + 101 I_B R_E = 11.3 \qquad (iii)$$

Similarly

$$V_{CC} = \beta I_B R_L + V_{CEact} + (1+\beta) I_B R_E$$

$$12 = 100\times 3\times 10^3 I_B + 0.3 + (1+100) I_B R_E$$

$$3\times 10^5 I_B + 101 I_B R_E = 11.7 \qquad (iv)$$

Solving the simultaneous equations (iii) and (iv) we get;

$$I_B = 1.9\,\mu\text{A}$$

$$R_E = 58\,\text{k}\Omega$$

$\therefore$ The value of $R_E = 58\text{ k}\Omega$

Exercise 3.3

1. What is a transistor? Describe various types of bipolar junction transistors.
2. Describe the working principle of npn and pnp junction transistors.
3. With the help of a circuit diagram describe the working principle of a transistor as an amplifier. List any applications of a common base amplifier.
4. Using the suitable diagram explain how a transistor operates as a switch. List any advantages of a transistor as a switch.
5. Figure 3.36 shows a circuit for a junction transistor voltage amplifier;
   a) What is the function of capacitors C_1 and C_2 and resistors R_1 and R_L?
   b) Derive expressions for the voltage gain and power gain.
   c) Given the supply voltage, V_{CC} = 10 V, V_{CE} = 0.31 V, collector current, I_C is 3 mA; assume that the base-emitter voltage, V_{BE} is 0.7 V and the current gain β is 100. Calculate the values of resistors R_L and R_1.

6. For a common emitter amplifier, current gain is equal to 50. If the emitter current is 6.6 mA, calculate the collector and base current. Also calculate the current gain when the emitter is working as a common base amplifier. If the base current is 100 μA and collector current is 3 mA, calculate the values of I_E, and β.
7. (a) Explain the formation of "depletion" layer and barrier "potential" in pn junctions.
   (b) With the help of a labelled circuit diagram, explain the use of a pn junction diode as a full wave rectifier. Draw the input and output waveforms.
8. (a) Draw a circuit diagram of npn transistor with its emitter–base junction at forward bias and base-collector junction at reverse bias. Describe briefly how it works.
   (b) Explain how a transistor in active state exhibits a low resistance at its emitter-base junction and high resistance at its base-collector junction.

3.4 Logic gates

A logic gate is a digital circuit that follows certain *logical* relationships between the input and the output signals. They are commonly known as *logic gates* because gates control the flow of information signals. Therefore *logic gates* are the basic building blocks of digital electronic devices which process digital signals in a specific manner. Digital signals can either be represented by logic high (1) or logic low (0).

One application of logic gates is as a switch, for example, when a switch is open, no current flows, hence the switch is in logic low (0) state, and when the switch is closed, current flows, hence the switch is logic high (1) state. In this subtopic the functions of each common logic gate are defined by a *truth table* that shows all the possibilities of the input logic level combinations and their respective output logic levels.

The truth tables and mathematical operations can help to understand the behaviour of logic gates. Mathematical operations of digital systems is carried out by the use of Boolean algebra. This is the mathematical operation analysis for handling variables that use binary values, either high (1) or low (0), sometimes referred to as ON or OFF states respectively. Logic gates are used in calculators, computers, digital watches, industrial control systems, in telecommunications and robots. Logic gates are decision making electronic switching circuits whose outputs can be expressed in the form of truth tables depending on the combination of their inputs.

3.4.1 Boolean algebra

In Boolean algebra there are three commonly used terms, *variable*, *complement* and *literal*. *Variable* means the symbol, normally an uppercase letter representing a logical quantity. This variable is either 0 or 1. *Complement* is changing of variable from 0 to 1 or from 1 to 0. It is indicated by the bar over the variable.

For example, the complement of variable is and read as complement or Not A or A bar. If, A = 1, $\overline{A}=0$ and if, A = 0, $\overline{A}=1$. *Literal* is either a variable or a complement. Different variables can be connected using mathematical operators to form Boolean expressions. The common Boolean operators are AND represented as multiplication ($\times$) or dot (.) OR is represented by plus (+) and NOT as the complement.

(a) Rules of Boolean algebra

Rules of Boolean algebra are very useful in simplifying Boolean expressions with the aim of reducing the size of a logic circuit. The following are some of the rules of Boolean algebra.

1. $X+0=X$
2. $X+1=1$
3. $X+X=X$
4. $X+\overline{X}=1$
5. $X.0=0$
6. $X.1=X$
7. $X.X=X$
8. $X.\overline{X}=0$
9. $\overline{\overline{X}}=X$

(b) Laws of Boolean algebra

In Boolean algebra, important laws are commonly used to simplify logical statements by making Boolean expressions to end up with simple or economic logic circuits. These include commutative, associative and distributive laws.

Commutative law

This law is applied to logical addition (OR) and logical multiplication (AND) of variables; it states that, "changing the order of variables when performing OR operation or AND operation does not affect final result". Mathematically is given:

$$A.B=B.A$$

$$A+B=B+A$$

Associative law

This law applies to logical addition (OR) and logical multiplication (AND) which states that, "when performing OR operation or AND operation of more than two variables, the order of grouping them makes no difference to final result". Mathematically is presented as:

$$A+(B+C)=(A+B)+C$$

$$A.(B.C)=(A.B).C$$

Distributive law

This law states that, "when performing OR operation of two or more variables followed by AND operation with a single variable gives the same result as performing AND operation of single variable with each of the two or more variables followed by OR operation". This enables the common variable to be factored out and makes simple expression. Mathematically is given by:

$$(A+B).C=A.C+B.C$$

Similarly, the law can be stated as

$$A+B.C=(A+B).(A+C)$$

Example 3.5

Use the rules of Boolean algebra to prove the following.

(a) $A+A.B=A$

(b) $A+\overline{A}.B=A+B$

(c) $(A+B).(A+C)=A+B.C$

Solution

(a) $A+A.B=A$

This law can be proved as follows:

$A+A.B=A.(1+B)$	Distributive law
$=A.1$	$1+A=1$
$=A$	$A.1=A$

(b) $A+\overline{A}.B = A+B$

The law can be proved as:

$A+\overline{A}.B = (A+\overline{A}).(A+B)$ Distributive law

$= 1.(A+B)$ $\quad A+\overline{A}=1$

$= A+B$ $\quad A.1 = A$

(c) $(A+B).(A+C) = A+B.C$

The law can be proved as follows:

$(A+B).(A+C) = A.A + A.C. + A.B + B.C$

$A.A + A.C. + A.B + B.C$ $\quad A.A = A$

$= A + A.C + A.B + B.C$ $\quad A + A.C = A$

$= A + A.B + B.C$ $\quad A + A.B = A$

$= A + B.C$

De Morgan's theorems

De Morgan provides two theorems which are very useful in digital circuit operations using Boolean algebra. One of the theorem states that, "the complement of the product is equal to the sum of the complement of each variable". Mathematically is given by:

$$\overline{A.B} = \overline{A} + \overline{B}$$

Another theorem states that, "the complement of the sum is equal to the product of the complements of each variable." Mathematically is represented by:

$$\overline{A+B} = \overline{A}.\overline{B}$$

Consider three variables function, De Morgan's theorems can be represented as;

$$\overline{A.B.C} = \overline{A} + \overline{B} + \overline{C}$$

$$\overline{A+B+C} = \overline{A}.\overline{B}.\overline{C}$$

3.4.2 Logic gates

A logic gate or a *gate* is a digital electronic circuit which contains one or more input terminals but one output terminal. The input terminals of the gate are represented to the left of the gate symbol and the output to the right. The logic gates are packed in an integrated circuits (IC) which may contain two or more gates as shown in Figure 3.41.

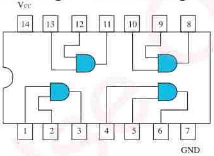


Figure 3.41 *AND gates in integrated circuit*

Logic gates may employ diodes and transistor to perform switching action. The common logic gates are NOT, AND, OR, NAND, NOR, XOR and XNOR and their operations are described as follows:

NOT gate

The NOT gate is an electronic circuit whose output is the inverted version of its input. It has only one input and one output. Sometimes it is known as INVERTER. Assume, the input variable is A which can either be 0 or 1, then the output will be A or $\overline{A}$ which is read as NOT A. The circuit symbol of the NOT gate is depicted in Figure 3.42.

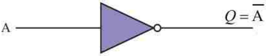


Figure 3.42 *NOT gate symbol*

The truth table of the NOT gate is presented in Table 3.2.

Table 3.2 Truth table of NOT gate

A	$Q=\overline{A}$
0	1
1	0

AND gate

The AND gate is an electronic circuit whose output is logic level 1 (HIGH) if and only if all the inputs are HIGH, otherwise the output is logic level 0 (LOW). This type of gate can have two or more inputs and the output is a logic multiplication of the inputs. Consider two inputs AND gate with variables A and B then the output can be expressed as A·B or AB which is read as A AND B. The logic symbol of the AND gate is shown in Figure 3.43.

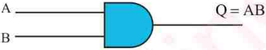


Figure 3.43 *AND gate symbol*

The truth Table of the AND gate is presented in Table 3.3.

Table 3.3 Truth table of AND gate

A	B	$Q=AB$
0	0	0
0	1	0
1	0	0
1	1	1

OR gate

The OR gate is an electronic circuit in which its output is logic level 0 (LOW) when all inputs are LOW, otherwise the output is logic level 1 (HIGH). This gate can have two or more inputs and its output is a logic sum of the inputs. Consider two inputs OR gate with variables A and B then the output is $A+B$ which is read as A OR B. Figure 3.44 present the logic symbol of the OR gate.

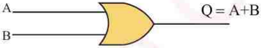


Figure 3.44 *OR gate symbol*

The truth table of the OR gate is depicted in Table 3.4.

Table 3.4 Truth table of OR gate

A	B	$Q=A+B$
0	0	0
0	1	1
1	0	1
1	1	1

NAND gate

It is noted that, the NAND gate is equivalent to AND gate followed by NOT gate. NAND gate is an electronic circuit in which its output is logic level 0 (LOW) if all the inputs are HIGH, otherwise the output is logic level 1 (HIGH). This gate can have two or more inputs and the output is the negation of the AND output. Consider a two inputs NAND gate with variables A and B, the output is expressed as $\overline{AB}$. Figure 3.45 shows the logic symbol of NAND gate and its equivalent gates.

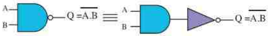


Figure 3.45 NAND gate symbol and its equivalent gates

Table 3.5 present the truth table of the NAND gate.

Table 3.5 Truth table of NAND gate

A	B	$Q=\overline{AB}$
0	0	1
0	1	1
1	0	1
1	1	0

NOR gate

Thus, NOR gate is equivalent to OR gate followed by NOT gate. NOR gate is an electronic circuit in which its output is HIGH (logic level 1) if all the inputs are LOW, otherwise the output is LOW (logic level 0). The NOR gate can have two or more inputs and the output is equivalent to the negation of the OR output. Consider two inputs of a NOR gate with variables A and B, the output can be expressed as $\overline{A+B}$. The logic symbol of the NOR gate and its equivalent gates are presented in Figure 3.46.

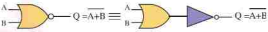


Figure 3.46 NOR gate symbol and its equivalent gates

Table 3.6. displayed the truth Table of the NOR gate.

Table 3.6 Truth table of NOR gate

A	B	$Q=\overline{A+B}$
0	0	1
0	1	0
1	0	0
1	1	0

Exclusive OR gate

Exclusive OR gate is sometimes denoted as EOR gate or as X-OR gate. Exclusive OR gate is a logical circuit in which its output is HIGH (logic level 1) when both of its inputs are at different logic level either LOW or HIGH, otherwise the output is LOW (logic level 0). Consider a two inputs XOR gate with variables A and B, the output can be expressed as $A\oplus B$ or $\overline{A}B+A\overline{B}$.
Figure 3.47 represents the logic symbol and Table 3.7 shows the truth table of the X-OR gate.

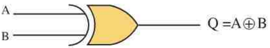


Figure 3.47 X-OR gate symbol

The truth Table of the X-OR gate is displayed in Table 3.7.

Table 3.7 Truth table of X-OR gate

A	B	$Q=A\oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

Exclusive NOR gate

Exclusive NOR gate is a logic circuit in which its output is LOW (logic level 0) if one of the inputs are either LOW or HIGH but not both, otherwise the output is HIGH (logic level 1). Exclusive NOR gate is sometimes denoted as E-NOR gate or as X-NOR gate. The output of X-NOR gate is the negation of the X-OR gate. It is noted that, X-NOR gate is equivalent to X-OR gate followed by NOT gate. Consider a two inputs X-NOR can gate with variables A and B, the output be expressed as $\overline{A \oplus B}$. Figure 3.48 shows the logic symbol of the X-NOR gate.

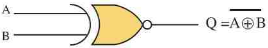


Figure 3.48 X-NOR gate symbol

The truth table of the X-NOR gate is presented in Table 3.8.

Table 3.8 Truth table of X-NOR gate

A	B	$Q = A \oplus B$
0	0	1
0	1	0
1	0	0
1	1	1

Example 3.6

For the gate combination shown in Figure 3.49, produce a truth table and draw its equivalent logic gate.

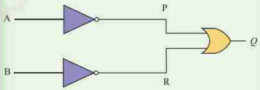


Figure 3.49 Logic gate circuit

Solution

Table 3.9 Truth table for the gate combination

Inputs		Outputs		
A	B	P	R	Q
0	0	1	1	1
0	1	1	0	1
1	0	0	1	1
1	1	0	0	0

The equivalent logic gate from the truth table is a NAND gate as shown in Figure 3.50.

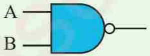


Figure 3.50 An equivalent logic gate

Example 3.7

An alarm system is controlled by three sensors, A, B and C. The system sounds its alarm when sensors A and B are on (i.e. logic level high) or when sensors B and C are on (i.e. logic level high). Draw the truth table for the system and hence design its logic circuit.

Solution

Boolean expression from the argument;

If variable are A, B and C

$Q(A,B,C) = A.B + B.C$

Introduce $A + \overline{A}$ and $C + \overline{C}$

$Q(A,B,C) = (A \cdot B)(C + \overline{C}) + (B \cdot C)(A + \overline{A})$

$Q(A,B,C) = ABC + AB\overline{C} + BCA + BC\overline{A}$

$Q(A,BC) = ABC + AB\overline{C} + \overline{A}BC$

where $ABC = BCA$

Table 3.10 A truth table

inputs			output
A	B	C	*Q(A,B,C)*
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	1

From $Q(A,B,C) = A\cdot B + B\cdot C$ the logic circuit is represented as shown in Figure 3.51.

Logic circuit designing. Write a Boolean expression from the truth table. Consider the input variables for high output only

$$Q(A,B,C) = ABC + AB\overline{C} + \overline{A}BC$$

$$Q(A,B,C) = AB(C+\overline{C}) + \overline{A}BC$$

$$Q(A,B,C) = AB + \overline{A}BC$$

$C + \overline{C} = 1$ Identity law

$$Q(A,B,C) = B(A + \overline{A}C) = B(A+C)$$

$A + \overline{A}C = A + C$ Distributive law

$$Q(A,B,C) = AB + BC$$

Figure 3.51 An equivalent logic gates circuit

Example 3.8

Use Boolean algebra to simplify the circuit in Figure 3.52 and implement the result using a minimum number of logic gates.

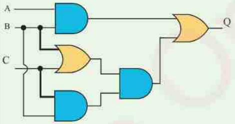


Figure 3.52 Logic gates circuit

Solution

Write a Boolean expression for this circuit by writing sub-expressions at the output of each gate corresponding to the respective input signals for each stage as shown below.

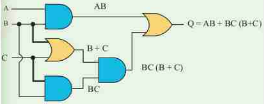


Figure 3.53 An equivalent logic gates circuit

Apply the rules of Boolean algebra to reduce the expression to its simplest form;

$$Q(A,B,C) = A.B + B.C(B+C)$$

$$Q(A,B,C) = A.B + B.B.C + B.C.C$$

Applying identity $A.A = A$

$$Q(A,B,C) = A.B + B.C + B.C$$

Applying identity $A + A = A$

$$Q(A,B,C) = A.B + B.C$$

Factorizing,

$$Q(A,B,C) = B.(A+C)$$

Therefore the circuit contains a minimum number of logic gates is:

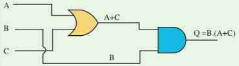


Figure 3.54 An equivalent logic gates circuit

Exercise 3.4

1. Define the term logic gate. Briefly discuss at least three types of logic gates.
2. Differentiate between:
   (i) AND gate and NAND gate;
   (ii) exclusive OR gate; and exclusive NOR gate.
3. Two car garages have a common gate which needs to open automatically when a car is detected to enter either of the garages. Draw a truth table and its equivalent logic gate circuit.
4. Simplify the following expressions using laws and rules of Boolean algebra and implement the results using a logic gate circuit.

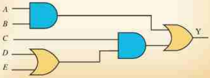
   (a) $Q = XY + \bar{X}Z + YZ$

   (b) $Q = X(Y+Z) + Y(Y+Z) + XY$

   (c) $Q = (X+Y)(X+Y+Z)$

   (d) $Q = XY + XYZ + XY\bar{Z}$

   (e) $Q = X\bar{Y}\bar{Z} + X\bar{Y}.$
5. Write down the Boolean expressions from Figure 3.55 of *Y(A,B,C,D,E)*.



Figures 3.55 Logic gates circuit

3.5 Operational amplifiers

Operational amplifiers (op-amps) are the basic transistor building blocks ofanalogue electronic circuits. These are voltage amplifying devices designed to be used with external feedback components, such as resistors and capacitors between their output and input terminals. These feedback components determine the function or "operation" of an amplifier. By virtue of the different feedback configurations, op-amps are used extensively in signal conditioning and filtering. They are also used to perform mathematical operations such as addition, subtraction, multiplication, integration, and differentiation, hence the name operational amplifier. Op-amps are widely used in computers, video and communication electronics. Op-amps are used in both digital and analogue electronics for the purpose of both *a.c.* and *d.c.* voltage gains.

In an op-amp the voltage supply is applied directly whereas in a transistor it is applied through a resistor and this counts to be its advantage. An op-amp is basically a three-terminal device which consists of two inputs of extremely high impedance, one called an *inverting input*, marked with a negative sign (–) and the other one called the *non-inverting input*, marked with a positive sign (+). This means that when a signal is applied to the non- inverting input it produces no phase shift in the output and when a signal is applied to the inverting input it produces 180^0 phase shift in the output(out of phase output). The third terminal represents an extremely low output impedance of the operational amplifier. There is an additional pair of power supply inputs giving equal positive and negative voltages i.e. $+V_s$ and $-V_s$ as shown in Figure 3.56.

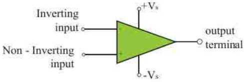


Figure 3.56 Circuit symbol for an Op-Amp

Properties of an operational amplifier:

(a) A very high open loop voltage gain A_o up to 10^5 for *d.c.* (typically 10^5 to 10^6).

(b) Extremely high input impedance, so that it draws virtually no current from the input signal.

(c) A very low output impedance.

(d) Very wide band width that should amplify almost any frequency from zero to very high frequencies.

(e) Very large Common Mode Rejection Ratio (CMRR) which is a metric used to quantify the ability of the device to reject common ratio mode signal that appear simultaneously and in phase on both inputs.

3.5.1 Open loop gain of an Op-Amps

The equivalent circuit of an op-amp is shown Figure 3.57. Thus, an op-amp can be regarded as being a device which generates voltage V_O. This output is proportional to the difference between the non- inverting input signal V_P and inverting input signal V_N. Mathematically is given by:

$$V_0 = A_0\left(V_P - V_N\right) \tag{3.24}$$

where A_0 is the proportionality constant known as *open loop voltage gain* of the amplifier.

The following situations may arise depending on the magnitude of V_N and V_P signals:

(i) V_0 will be positive when V_P is greater than V_N

(ii) V_0 will be negative when V_N is greater than V_P

(iii) V_0 will be zero when V_P is equal to V_N

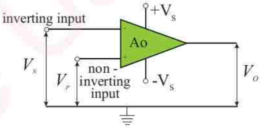


Figure 3.57 A circuit of an op-amp

The open loop voltage gain (A_0) is related to the output voltage V_{out} and the input voltage V_{in} as:

$$A_0 = \frac{V_0}{\left(V_P - V_N\right)} = \frac{V_0}{V_{in}} \text{ whereby } V_{in} = V_P - V_N$$

$$A_0 = \frac{V_0}{V_{in}} \tag{3.25}$$

Example 3.9

The open loop voltage gain of an operational amplifier Figure 3.58 is 10^5 and the supply voltage is ± 9V.

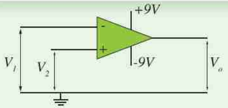


Figure 3.58 The operational amplifier

a) Calculate the minimum potential difference which may be applied between the inputs terminals to produce saturation of amplifier.

(b) In the circuit above V_1 and V_2 are input voltages to an operational amplifier of large open loop voltage gain, V_1 is a sinusoidal signals of peak value 6 V and V_2 is 3 V *d.c.* On the same axes sketch graphs of voltage against time for each of the two inputs voltages and for the output voltage.

Solution

(a) The input voltage$=\dfrac{\text{output voltage}}{\text{voltage gain}}$

At saturation the output voltage is equal to supply voltage.

$$\text{input voltage}=\frac{9\text{V}}{10^5}=9.0\times10^{-5}\text{V}=90\propto\text{V}$$

(b) Figure 3.59 shows the sketch graphs of voltage against time plotted on the same axes for each of the two input voltages and for the output voltage.

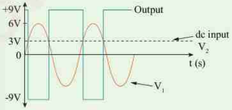


Figure 3.59 A sketch of voltages against time

Transfer characteristic of operational amplifiers

The amplification is achievable provided the output voltage, V_o is less than that of the supply voltage V_S. When V_o reaches V_S the op-amp is said to be saturated and the outputs lies between $+V_S$ and $-V_s$ as shown in Figure 3.60. The region between A and B in the graph is linear, and this region is where op-amp is operated as an amplifier.

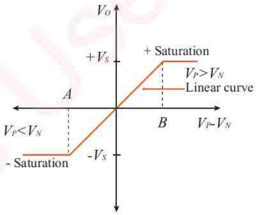


Figure 3.60 Transfer characteristic of op-amp

Open loop voltage gain varies with frequency as shown in Figure 3.61. As the frequency of the input signal increases, the open loop voltage gain falls. The limited linear behaviour is caused by the very high open loop voltage gain A_o, and the higher it is, the greater the limitation and vice versa. It should be noted that the output voltage should never exceed the supply voltage.

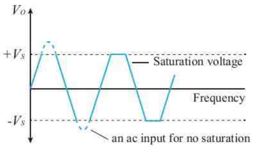


Figure 3.61 Frequency response in open loop voltage gain

3.5.2 Feedback in the operational amplifier

Feedback in op-amp is the process of adding a fraction of the output signal back to the input, as shown in Figure 3.62. Feedback depends upon whether the fed back signal aids or opposes the input signal. There are two types of op-amp feedback namely *negative* and *positive* feedback.

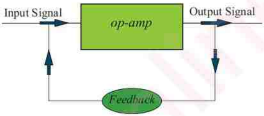


Figure 3.62 Block diagram for a feedback in an op-amp

When the signals at the input and the output are out of phase, the feedback signal is said to be negative. A fraction of the output which is fed back to the inverting input of an op-amp has an effect of subtracting this feedback signal from the input signal so that it reduces the overall gain of an amplifier. As a result, the gain is highly stable and is constant over a large range of input voltage and frequencies. Therefore, the circuit has the advantage of having gain which is independent of the characteristics of an amplifier itself. When the input and the output signals are in phase, the feedback signal is said to be positive. In positive feedback, a fraction of the output is fed back to the non-inverting input. This type of the feedback tends to increase the gain of the op-amp; however it has the disadvantages of increased distortion and instability. Therefore positive feedback is rarely employed in amplifiers. One important use of positive feedback is that, it is employed in relaxation oscillators which generate alternating current signals from a *d.c.* source.

3.5.3 Operational amplifier circuits

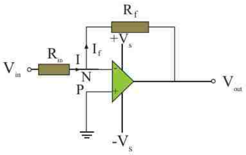
Operational amplifiers can be connected in different circuits to provide various operating characteristics. In this subsection most common circuit connections will be dealt with.

(a) Inverting voltage amplifier

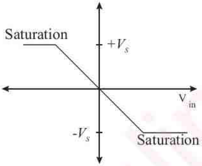
When an input signal is applied to the inverting input of an op-amp and non-inverting input is earthed (grounded) as shown in Figure 3.63 (a), the resulting circuit is called an inverting amplifier. The fraction of the output is fed back to the inverting input so that the value of the input voltage V_{in} is reduced.

The negative feedback is provided by the feedback resistor (R_f). If point P at the non-inverting input is grounded, then point N at the inverting input is said to be a virtual earth due to the concept of infinite input impedance and very high open loop gain making potential difference between the two inputs zero i.e.

Voltage gain is $\frac{V_0}{V_{in}}$ for infinite gain as $V_{in} \rightarrow 0$. Figure 3.63 (b) shows the variation of the output with input voltages. This variation is known as transfer characteristics of an inverting amplifier.



(a)



(b)

Figure 3.63 The inverting amplifier circuit and its transfer characteristics

The potential difference (*p.d.*) across the resistor (R_{in}) is therefore $V_{in} - V_N$ and the *p.d.* across the resistor R_f is $V_N - V_0$. At the junction N, the current I is given as:

$$I = I_f$$

$$I = \frac{V_{in} - V_N}{R_{in}} \text{ and } I_f = \frac{V_N - V_0}{R_f}$$

Since $V_P = 0$ (grounded),

$$V_P = V_N = 0 \text{ (virtual earth)}$$

It follows that $\frac{V_{in}}{R_{in}} = -\frac{V_0}{R_f}$

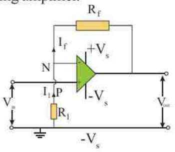
Therefore, the closed loop voltage gain (A_{CL}) for the inverting voltage amplifier with a feedback resistor R_f and input resistor R_{in} is given as:

$$A_{CL} = \frac{V_o}{V_{in}} = -\frac{R_f}{R_{in}} \qquad (3.26)$$

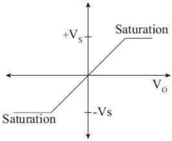
The closed loop voltage gain (A_{CL}) depends only on the two resistors.

(b) Non- inverting voltage amplifier

In this amplifier, the input voltage is applied to the non–inverting terminal, and the fraction of the output is fed back to the inverting input as shown in Figure 3.64 (a). This gives an output voltage that is in phase with the input voltage. Figure 3.64 (b) shows the transfer characteristics of non-inverting amplifier.



(a)



(b)

Figure 3.64 The non- inverting amplifier circuit and its transfer characteristics

A non-inverting amplifier is a special case of the differential amplifier in which a circuit's inverting input is grounded and the input voltage V_{in} is applied to the non-inverting input.

By using the virtual ground concept at point N, one can calculate the current through resistor R_1 and the same current must be passing through R_f.

At the junction N:

$$I_1 = I_f$$

$$I_1 = \frac{0 - V_N}{R_1} \text{ and } I_f = \frac{V_N - V_0}{R_f}$$

Since $V_P = V_{in} = V_N$

It follows that $-\frac{V_{in}}{R_1} = \frac{V_{in} - V_o}{R_f}$

$$\frac{V_o}{R_f} = \frac{V_{in}}{R_f} + \frac{V_{in}}{R_1} = V_{in}\left(\frac{1}{R_f} + \frac{1}{R_1}\right)$$

$$A_{CL} = \frac{V_o}{V_{in}} = 1 + \frac{R_f}{R_1}$$

It can also be done by using potential divider theorem, as:

$$V_N = \left(\frac{R_1}{R_f + R_1}\right)V_o$$

Since $V_{in} = V_P$

and $V_{in} = V_P = V_N$

Then $V_{in} = \left(\frac{R_1}{R_f + R_1}\right)V_O$

Thus, $A_{CL} = \frac{V_o}{V_{in}} = 1 + \frac{R_f}{R_1}$ *(3.27)*

Therefore, the closed loop voltage gain (A_{CL}) for the non-inverting voltage amplifier with a feedback resistor R_f and input resistor R_{in} can be determined by using equation (3.27). The closed loop voltage gain (A_{CL}) also depends only on the two resistors.

Example 3.10

(a) Find the closed loop voltage gain in Figure 3.65.

(b) What is the name of this operational amplifier circuit?

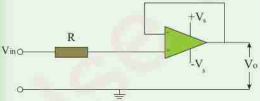


Figure 3.65: A special case of non-inverting op-amp

Solution

(a) From equation 3.27

$$A_{CL} = \frac{V_o}{V_{in}} = \left(1 + \frac{R_f}{R_1}\right)$$

but $R_f = 0$ and $R_1 = R$

Hence:

$$A_{CL} = \frac{V_o}{V_{in}} = 1 + \frac{0}{R_1} = 1$$

The closed loop voltage gain is a unit.

(b) This operational amplifier is known as Voltage follower since $V_{in} = V_O$.

It is noted that the most important features of voltage follower configuration is that it has high input impedance and extremely low output impedance. These features make it a nearly ideal buffer amplifier to be connected between high impedance sources and low impedance loads.

Example 3.11

The diagram in Figure 3.66 shows an operational amplifier circuit with feedback. If the operational amplifier has a gain of 1,000.

(i) What will V_{out} be when V_{in} = 0.5 V?

(ii) If $R_1 = 2\ k\Omega$, find the magnitude of R_2

(iii) State what the given amplifier circuit represents.

(iv) Give the reason for employing negative feedback in the circuit.

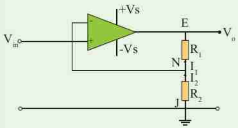


Figure 3.66 An operational amplifier with a negative feedback

Solution

(i) From the KCL at N node and the concept of virtual ground, then

$$I_1 = I_2$$

From the potential divider configuration:

$$V_N = \left(\frac{R_2}{R_1 + R_2}\right)V_o$$

$$V_N = V_{in} \text{ (virtual earth)}$$

$$V_{in} = \left(\frac{R_2}{R_1 + R_2}\right)V_o$$

$$V_0 = \left(\frac{R_2 + R_1}{R_2}\right)V_{in}$$

$$A = \frac{V_0}{V_{in}} = \left(\frac{R_2 + R_1}{R_2}\right)$$

Thus, $V_o = 1000 \times 0.5\ V = 500\ V$

(ii) If $R_1 = 2\ k\Omega$

$$A = \frac{V_0}{V_{in}} = \left(\frac{R_2 + R_1}{R_2}\right)$$

$$R_2 = \left(\frac{R_1}{A-1}\right) = \left(\frac{2000}{1000-1}\right) = 2.002\ \Omega$$

(iii) The given amplifier circuit represents a non-inverting voltage amplifier since the output is in phase with input.

(iv) The negative feedback improves stability of the gain by making it dependent on the values of relatively stable resistors, and not on the very temperature sensitive semiconductor in the operational amplifier.

(c) Summing amplifier

As its name suggests, summing amplifier can be used to combine the voltage present on multiple inputs into a single output voltage as shown in Figure 3.67.

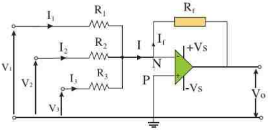


Figure 3.67 The summing amplifier circuit

Applying Kirchhoff's current law at point N, we get:

$$I_1 + I_2 + I_3 = I = I_f$$

$$\frac{V_1 - V_N}{R_1} + \frac{V_2 - V_N}{R_2} + \frac{V_3 - V_N}{R_3} = \frac{V_N - V_0}{R_f}$$

Since $V_P = 0$ (grounded), then:

$$V_P = V_N = 0 \text{ (Virtual earth)}$$

$$\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} = -\frac{V_o}{R_f}$$

$$V_0 = -\left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}\right) R_f$$

If $R_1 = R_2 = R_3 = R_{in}$ the equation reduces to $V_0 = -\frac{R_f}{R_{in}}\left(V_1 + V_2 + V_3\right)$

A direct voltage addition can be obtained when all the resistances are of equal value and R_f is equal to R_{in} i.e.

$(R_1 = R_2 = R_3 = R_{in} = R_f)$

$$V_0 = -\left(V_1 + V_2 + V_3\right)$$

If more inputs voltage signals are involved they can be added; each individual input finds its resistance, R_i as the only input impedance. This is because the input signals are effectively isolated from each other by the "virtual earth" node at the inverting input of an op-amp.

Generally,

$$V_0 = -R_f \sum_{i=1}^{n} \frac{V_i}{R_i} \quad i = 1,2,3.....n \qquad (3.28)$$

Equation (3.28) shows that the summing amplifier amplifies input signals and produce the output voltage signal which is proportional to the algebraic "sum" of the individual input voltage.

Example 3.12

Find the output voltage of the following summing amplifier circuit in Figure 3.68 if $R_1 = 1\ \text{k}\Omega$, $R_2 = 2\ \text{k}\Omega$, $R_f = 10\ \text{k}\Omega$, $V_1 = 2$ mV and $V_2 = 5$ mV

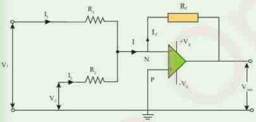


Figure 3.68 Summing amplifier with two inputs

Solution

$$V_0 = -\left(\frac{V_1}{R_1} + \frac{V_2}{R_2}\right) R_f$$

$$V_o = -\left(\frac{2\times10^{-3}}{1\times10^{3}} + \frac{5\times10^{-3}}{2\times10^{3}}\right)\times10\times10^{3}$$

$$= -4.5\times10^{-2}\ \text{V}$$

Then the output voltage of the summing amplifier circuit in Figure 3.68 is given as - 45 mV. It is negative because of input voltages V_1 and V_2 are out of phase by π radians.

(d) Operational amplifier as an integrator

As its name implies, the op-amp integrator is an operational amplifier circuit Figure 3.69 that performs the mathematical operation of integration. The op-amp integrator produces an output voltage which is proportional to the integral of the input voltage.

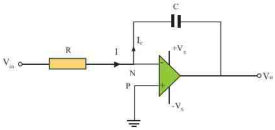


Figure 3.69 Voltage integrator circuit

Assuming that the input impedance of the op-amp is infinite (ideal op-amp), i.e. no current flows into the op-amp terminal.

Using Kirchhoff's current law at the junction N:

$$I = I_C$$

$$I = \frac{V_{in} - V_N}{R} \text{ and } I_C = \frac{dQ}{dt} \text{ but } Q = CV$$

$$\frac{V_{in} - V_N}{R} = C\frac{d}{dt}\left(V_N - V_0\right)$$

Since $V_P = 0$ (grounded), then;

$$V_P = V_N = 0 \text{ (Virtual earth)}$$

It follows that $\frac{V_{in}}{R} = -C\frac{d}{dt}V_0$

Applying an integration on both sides of this equation then;

$$V_0 = -\frac{1}{RC}\int V_{in}\, dt \qquad (3.29)$$

This equation shows that the output voltage is the integration of the input voltage (with an inversion) and a multiplier of $\frac{1}{RC}$.

Example 3.13

Refer to the circuit in Figure 3.69, describe what happen to V_o when V_{in} is raised suddenly from 0 to 1 V and remain at that voltage, if $C = 3\ \mu\text{F}$ and $R = 1.5\ \text{M}\Omega$.

Solution

Applying equation (3.29)

$$V_0 = -\frac{1}{RC}\int V_{in} dt$$

$$V_0 = -\frac{1}{1.5\times10^6\times3\times10^{-6}}\int V_{in} dt$$

$$= -0.22\int V_{in} dt$$

$$V_o = -0.22\int 1 dt = -0.22t$$

Negative means the V_o is inverted, and it varies with time hence the name ramp generator.

(e) Operational amplifier as a differentiator

As a differentiator, an op-amp is used to perform a mathematical operation of differentiation. The circuit which can achieve that operation is as illustrated in Figure 3.70.

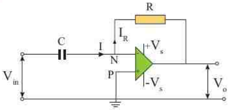


Figure 3.70 Voltage differentiator circuit

From Kirchhoff's current law at node N,

$$I = I_R$$

$$I_R = \frac{V_N - V_0}{R} \text{ and } I = \frac{dQ}{dt}$$

but $Q = CV$

$$\frac{V_N - V_0}{R} = C\frac{d}{dt}\left(V_{in} - V_N\right)$$

Since $V_P = 0$ (grounded), then:

$$V_P = V_N = 0 \text{ (Virtual earth)}$$

It follows that $V_o = -RC\dfrac{dV_{in}}{dt}$ $\quad$ (3.30)

A differentiator produces an output voltage that is proportional to the rate of change of the input voltage.

(f) Operational amplifier as a subtractor

An Op-amp subtractor circuit is designed to perform the mathematical operation of subtraction. The circuit that can implement this operation is as shown in Figure 3.71.

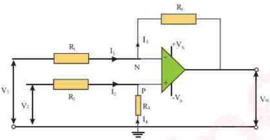


Figure 3.71 Voltage subtractor circuit

Assuming the input impedance is infinite, applying Kirchhoff's current law at N and P nodes,

$$I_1 = I_3 \text{ and } I_2 = -I_4$$

$$I_1 = \frac{V_1 - V_N}{R_1};\, I_3 = \frac{V_N - V_o}{R_3}$$

Then,

$$\frac{V_1 - V_N}{R_1} = \frac{V_N - V_o}{R_3}$$

$$V_N = \frac{R_3 V_1 + R_1 V_o}{R_1 + R_3} \qquad (i)$$

$$I_2 = \frac{V_2 - V_P}{R_2};\, I_4 = \frac{0 - V_P}{R_4}$$

$$\frac{V_2 - V_P}{R_2} = -\left(\frac{0 - V_P}{R_4}\right) = \frac{V_P}{R_4}$$

$$V_P = \left(\frac{R_4}{R_4 + R_2}\right)V_2 \qquad (ii)$$

Assuming $V_N = V_P$ (virtual earth) then:

$$\frac{R_3 V_1 + R_1 V_0}{R_1 + R_3} = \left(\frac{R_4}{R_4 + R_2}\right)V_2 \qquad (iii)$$

Manipulate equation (iii) you have:

$$V_0 = \left(\frac{R_1 + R_3}{R_1}\right)\left(\frac{R_4}{R_4 + R_2}\right)V_2 - \frac{R_3}{R_1}V_1 \qquad (iv)$$

Assume $R_1 = R_2 = R_3 = R_4 = R$ then equation (iv) becomes:

$$V_0 = V_2 - V_1 \qquad (3.31)$$

Hence the output voltage V_o, is a subtraction of the voltage at the non- inverting terminal V_2 and the inverting terminal, V_1.

(g) Operational amplifier as a voltage comparator

An op-amp as a comparator, it compares one analogue voltage level with another voltage level to produce an output voltage (V_0). The circuit does not use feedback as shown in Figure 3.72. The resistor R_1 and R_2 form a potential divider between the two voltage rails. The voltage V at the junction of the two resistors is the switching voltage for the circuit.

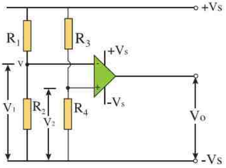


Figure 3.72 Voltage comparator circuit

The potential difference across the series combination of R_1 and R_2 is the same as that across R_3 and R_4. It therefore follows that, if $\frac{R_2}{R_1}$ is greater than $\frac{R_4}{R_3}$, and V_1 is greater than V_2, then the output voltage V_0 is negative, and vice versa.

The voltage comparator circuit can be used as a sensitive device for measuring the value of unknown resistance. Assuming R_1, R_2, R_3 and R_4 are the four resistances in a Wheatstone bridge circuit, V_0 is equal to zero only when $\frac{R_2}{R_1}$ equals to $\frac{R_4}{R_3}$. The circuits can also be used to switch on or off some lighting or heating system if it is connected to a light dependent resistor (LDR) or thermistor (Th), respectively.

Activity 3.8

Design a simple operational amplifier circuit which can be used to detect the temperature rise in a room.

Exercise 3.4

1. Give a brief description of an operational amplifier.
2. What is an open loop voltage gain as far as op-amps are concerned?
3. Discuss at least three main properties of operational amplifiers.
4. Briefly explain about transfer characteristics of operational amplifiers.
5. What is meant by the term feedback as applied in op-amps? Discuss different types of feedback.
6. Explain the concept of virtual earth. Derive an expression for the gain of non –inverting amplifier.
7. Using the circuit in Figure 3.71 and given $V_1 = 0.45\,V$, $V_2 = 1.5$ V and $R_1 = R_2 = R_3 = R_4 = 95$ kΩ
   (i) Write down the relationship between output voltage V and input voltages V_1 and V_2.
   (ii) What will be the value of the output voltage?
8. Consider the circuit of operational amplifier in Figure 3.69 where $R = 3$ kΩ, $C = 3\,\mu$F. Find the output voltage when:
   (i) the input voltage signal $V_i = 2\sin\omega t$ and $f = \frac{20}{\pi}$ Hz;
   (ii) the input voltage signal varies from 0 V to 9 V in 5 s. (Assume the input voltage changes at constant rate).

3.6 Telecommunication

Communication is the exchange of information from one point to another through a series of processes. The information in this case may be pictures, videos, sounds or text. When the exchange of information takes place over a distance, this defines telecommunication. Hence *telecommunication* is the transmission or reception of signals in form of text images, sounds, or intelligence of any nature by wire, radio, optical, or any other electromagnetic system.

On the other hand telecommunication is simply communication at a distance. The communication technology uses channels to transmit information as electrical signals either through a physical medium such as optical fibre or in form of the electromagnetic waves.

The basic communication system consists of three main parts as illustrated in Figure 3.73.

These parts can be described as follows:

(a) A transmitter that takes information and converts it into a transmitted signal.
(b) A transmission medium also called "channel" that carries the signal.
(c) A receiver that takes the signal from the channel and converts it back into the information for the recipient.

Figure 3.73 Basic communication system

Transmitter: A transmitter is an electronic device that transforms message signal to suitable form that can be transmitted over communication channel.

Figure 3.74 shows an example of TV and radio transmitting dish. It consists of an input transducer that converts the information to its equivalent electrical or optical signal for transmission, and a modulator which converts the information signal into a form which can easily be transmitted and detected by the receiver. A *transducer* is a device that converts one form of energy into another form of energy, for instance, microphone coverts sound energy into electrical energy.

Figure 3.74 TV and radio transmitter dish

Communicating channel: This is a medium through which modulated signal is transmitted from transmitter to receiver. Communication channel may be free space (air) or physical channel, such as optical wave guide or fibres.

Receiver: A receiver is a device which receives the modulated signals from the communication channel, demodulates and amplifies them to their original form. It consists of a demodulator circuit, an amplifier and output transducer circuit, such as loud speaker which converts electrical energy into sound energy. Figure 3.75 shows an example of TV and radio receiver dish.

Figure 3.75 TV and radio receiver dish

Repeater: A repeater refers to a combination of a receiver and a transmitter. It picks up the signal from the transmitter, amplifies and retransmits it to the receiver sometimes with a change in carrier frequency. It is used to extend the range of a communication system as shown in Figure 3.76. For example, a communication satellite is a powerful repeater station in space.

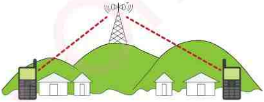


Figure 3.76 A repeater station

Basic modes of communication

The basic modes of communication include; *Point–to-point communication mode and broadcast mode*. In point-to-point communication mode, the message is transmitted over a link between a single transmitter and a single receiver e.g. conversation between two people through a telephone. In broadcast mode, there is a single transmitter and a large number of receivers e.g. radio broadcasting and television telecast.

Modulation

To convey an information signal over a distance through a channel, a *carrier signal* is required. This carrier signal must have the ability to travel from the transmitting point to the receiving point without attenuation (i.e. the loss of the strength of the signal while propagating through the medium). The question is: How should the information signal be added to the carrier signal? The solution is: Some characteristics of the carrier signal must change in accordance with the intensity of the information signal.

Modulation is defined as a process of changing some characteristics (e.g. amplitude, frequency and phase) of the carrier wave in accordance with the intensity of the information signal. If the amplitude of the carrier signal is modified by the intensity of the information signal, the process is known as ***amplitude modulation (AM)***. On the other hand, if the frequency of the carrier signal is modified by the intensity of the information signal, the process is known as ***frequency modulation (FM)***, and if the phase of the carrier signal is modified by the intensity of the information signal, the process is known as ***Phase modulation (PM).***

Amplitude Modulation (AM)

Amplitude Modulation (AM) is the modulation technique whereby the amplitude of the carrier signal is varied in accordance with the intensity of the information signal. However, the frequency of the carrier signal stays constant. The process of the amplitude modulation is shown in Figure 3.77.

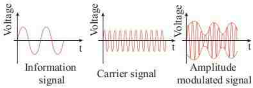


Figure 3.77 Amplitude Modulation

Modulation Index (M): An important consideration in amplitude modulation is the description of the extent to which the amplitude of the carrier wave is changed by the modulating signal.

This extent is described by a factor called *modulation index* (M) which is defined as the ratio of the change of amplitude of the carrier wave to the amplitude of the unmodulated carrier wave.

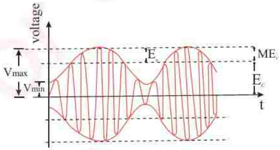


Figure 3.78 Modulation index

If the maximum and minimum voltage of AM wave are V_{max} and V_{min} respectively, it is clear from the Figure 3.78 that:

$$E_c = \frac{V_{max} + V_{min}}{2} \text{ and } E_s = \frac{V_{max} - V_{min}}{2}$$

$$M = \frac{E_s}{E_c} = \frac{\frac{V_{max} - V_{min}}{2}}{\frac{V_{max} + V_{min}}{2}} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

$$M = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \qquad (3.32)$$

where E_c is a carrier amplitude, E_s is a signal amplitude. It should be noted that the modulation index (M) generally lies between 0 and 1.

Bandwidth

The amplitude modulated signal may be regarded as being made up of the carrier signal of frequency f_c and other side frequencies. The process of modulation doesn't change the carrier frequency but it produces two new frequencies. If f_m is the frequency of the modulating signal then the lower frequency signal, $f_c - f_m$, and the upper frequency signal, $f_c + f_m$ can be observed. These signals (i.e. $f_c - f_m$ and $f_c + f_m$) are known as side frequencies. Practically, the modulating signal results into a band of frequencies called *lower and upper side bands* spreading below and above the carrier frequency (f_c) by the value of highest modulating frequency (f_m) respectively. The total range in frequency is called *bandwidth* of the signal, as shown in Figure 3.79.

Generally the bandwidth of an amplitude modulated signal is twice the highest frequency contained in the modulating signal. It should be noted that the tuned amplifier which is called upon to amplify the modulated wave must have the required bandwidth to include sideband frequencies. If the tuned amplifier has insufficient bandwidth, it may not be able to reproduce the original signal.

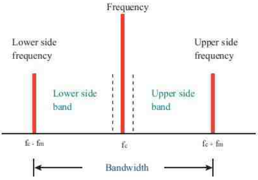


Figure 3.79 Modulated signal with side frequencies and side bands

Disadvantages of Amplitude Modulation.

i) It has high noise reception.
ii) It has low efficiency.
iii) Operation range is small.
iv) Lack of audio quality.

Despite of their disadvantages, amplitude modulation is useful in various cases.

Example 3.14

A sinusoidal carrier voltage of frequency 2 MHz and amplitude 70 V is amplitude modulated with sinusoidal voltage of frequency 4 kHz producing modulation factor of 55%. Determine:

i) the frequency of the lower side band;
ii) the frequency of the upper side band;
iii) the bandwidth of resultant; modulated signal; and
iv) the amplitude of upper and lower sideband.

Solution

Given that,

$f_c = 2\,MHz,\ f_m = 4\ kHz,\ M = 0.55$ and

$E_c = 70$ V

(i) Lower sideband $= f_c - f_m$

$= (2000\text{-}4)$ kHZ $=$ 1996 kHz

$=$ 1.996 MHz

(ii) Upper sideband $= f_c + f_m$

$= (2000+4)$ kHz $=$ 2004 kHz

$=$ 2.004 MHz

(iii) Bandwidth = upper sideband - lower sideband frequency

$= (2004\text{-}1996)$ kHz

$=$ 8 kHz

(iv) Amplitude $= \dfrac{ME_c}{2}$

$$= \frac{0.55 \times 70}{2} = 19.25\text{V}$$

Frequency Modulation (FM)

Frequency modulation (FM) is the technique whereby the frequency of the carrier signal is changed in accordance with the intensity of the information signal. In this case, the amplitude of the carrier signal stays constant as illustrated in Figure 3.80.

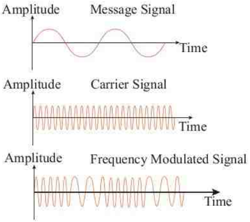


Figure 3.80 Frequency Modulation

Advantages of FM over AM

Frequency Modulation has many advantages in the modern electronic technology when compared with Amplitude Modulation. Some of these advantages are:

i) It has a very high signal-to-noise ratio.
ii) It gives a high fidelity reception, meaning that there is a very little distortion during transmission and a true copy of information is received; this is very useful in FM radio communication.
iii) The efficiency of transmission is very high.
iv) It has a large number of side bands, which makes it useful in stereo sound communication.

Phase Modulation (PM)

Phase modulation (PM) refers to modulation that encodes information as a variation in the instantaneous phase of a carrier wave. Phase Modulation is an integral part of many digital transmission coding schemes that underlie a wide range of technologies like satellite and television. It is also widely used for transmitting radio waves. The phase modulation characteristic is as shown in Figure 3.81.

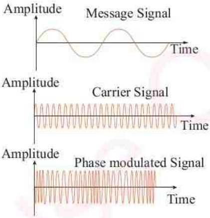


Figure 3.81 Phase modulation

Figure 3.82 is a representation of carrier wave modulation showing corresponding pattern of the three types of modulations which are Amplitude Modulation (AM), Frequency Modulation (FM) and Phase Modulation (PM).

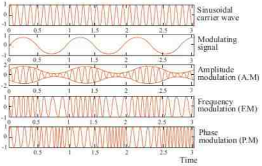


Figure 3.82 Modulation of a carrier wave

Radio and television signals transmission and reception

In a broadcast system, the central high-powered broadcast tower transmits a high frequency electromagnetic wave to numerous low-powered receivers. The high frequency wave sent by the tower is modulated with a signal containing visual, audio or a combination of two media information. The receiver is then tuned to pick up the high-frequency wave. A demodulator is used to retrieve the signal containing the visual, audio or combination of the two media information.

Factors contributing to transmission

Transmission of the low frequencies such as audio signals which have a frequency range between 20 Hz and 20 kHz over a long distance directly (i.e. without modulation), factors which may prevent the transmission and method to overcome them are discussed underneath.

Size of antenna or aerial

For transmitting a signal an antenna is required. This antenna should have a length (l) equivalent to one fourth of the signal wavelength (λ), so that the antenna properly matches the time variation of the signal. Mathematically;

$$l = \frac{\lambda}{4} \qquad (3.33)$$

For an electromagnetic wave of frequency 15 kHz, the wavelength (λ) is 20 km and an antenna of length 5 km result. This antenna is very large which is practically impossible to construct and operate. Hence direct transmission of such signals is not possible. Transmission with reasonable antenna lengths is necessary if transmission frequency is high. Also for a frequency of 1 MHz, the wavelength is 300 m and an antennae of length 75 m would be required; again this is not practical. Therefore, there is a need to modulate the information contained in our original low frequency baseband signal into high frequencies before transmission.

Effective power radiated by an antenna

A theoretical study of radiation from a linear antenna (length, l) shows that the radiated power P is proportional to the square of l and also inverse proportional to the square of wavelength. Mathematically:

$$P \propto \left(\frac{l}{\lambda}\right)^2 \qquad (3.34)$$

This implies that for the same antenna length, the power radiated increases with decreased λ. Hence, the effective power radiated by a long wavelength e.g. 20 km in audio signal would be small. For a good transmission, high powers are necessary and hence this also points to the need of using high frequency transmission.

Mixing up of signals from different transmitters

Transmission of baseband signals directly is more *practical* in nature. Suppose many transmitters are transmitting baseband information signals simultaneously. These signals will get mixed up and there is no simple way to discriminate them. Therefore a possible solution is by using communication at high frequencies and assigning a *band* of frequencies to each message signal for its transmission.

Forms of transmission signals

The format in which information or data is transmitted is either analogue or digital. Consequently, in analogue modulation a higher frequency signal is generated by varying some characteristic of a high frequency signal (carrier) on a continuous basis i.e. an infinite number of baseband signals, as illustrated in Figure 3.83(a).

In digital modulation, signals are converted to binary data, encoded as a set of discrete values and translated to higher frequency, as illustrated in Figure 3.83 (b). Digital modulation is more complex, but reduces the effect of noise and finite number of baseband signals.

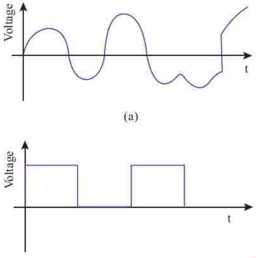


Figure 3.83 Analog and digital signals

Moving from analog to digital broadcasts can be made possible by the production of cheaper, faster and more capable integrated circuits (*IC*). The main advantage of digital processing and transmission is the elimination of problems such as snowy pictures, ghosting and other distortions identified with analogue broadcasting. These problems occur because of the nature of analogue transmission, that is, the perturbations due to noise from unwanted signal will be evident in the final output. Digital transmission overcomes this problem because digital signals are reduced to discrete values when received and thus small perturbations do not affect the final output.

Note *In a simplified example, if a binary message 1011 was transmitted with signal amplitudes [1.0 0.0 1.0 1.0] and received with signal amplitudes [0.9 0.2 1.1 0.9] it would still decode to the binary message 1011 which is a perfect reproduction of what was sent. From this example, a problem with digital transmissions can also be seen, in that, if the noise is high enough, it can significantly alter the decoded message. Using forward error correction a receiver can correct a handful of bit errors in the resulting message but too much noise will lead to incomprehensible output and breakdown of the transmission.*

Radio transmission

For the massage or information to be transmitted from place to place a transmitter is necessary. The major function of a transmitter is to transform the information into an appropriate form and transmit it into free space (air) through a transmitting antenna as illustrated in the block diagram of a transmitter for radio broadcasting, Figure 3.84.

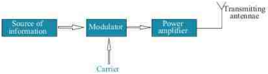


Figure 3.84 Block diagram for transmission system

The information from microphone, camera, computer, and TV are represented as electrical signals, these signals can be sent from place to place using either radio waves, copper cables or optical fibers (wave guide).

In radio waves, the electrical signal is radiated from the aerial in the form of electromagnetic waves. These waves travel outwards from the aerial with the speed of light. Electromagnetic waves in the frequency range 30 kHz to 300 GHz are generally referred to as radio waves. The optical fibre Figure 3.85 is used in transmission of electrical signal, this comprise of a very fine strand of pure glass, surrounded by protective covering. The glass fibre itself is thinner than a human hair. The pulses of light or infrared radiation travel along the fibre as a result of total internal reflection. These pulses carry an information signal along the fibre.

Figure 3.85 Optical fibre

Audio frequency signals may be transmitted directly by cable; however, in radio and TV, a carrier wave is needed. It has a constant amplitude and its waveform is sinusoidal. The carrier wave has higher frequency than information signal.

Signals from an information source are added to the carrier in the modulator by the process of modulation and the circuit that performs this function is called a modulator. The modulated signal is amplified before being sent to the transmitting antenna which converts the electrical signal into radio waves and transmit them into space.

Radio reception

Figure 3.86 shows a block diagram of the receiver for radio broadcasting. The transmitted signal i.e. modulated radio wave is captured by the receiver through the receiving antenna. This antenna captures the radio waves from different transmitting stations. The receiving antenna is therefore to be followed by a tuned amplifier to amplify the desired signal. In addition, to enable further processing of tuned amplified signal, the carrier frequency is commonly changed to a lower frequency in intermediate frequency stage. The output from intermediate frequency stage is fed to the demodulator (detector) which separate the audio signal from the modulated signal, thus demodulation (detection) process. The demodulated signal is not strong enough to be made use of, and hence it is required to be amplified by the audio amplifier and then fed to the output transducer (loud speaker) for reproduction into sound signals.

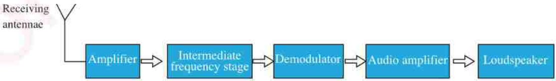


Figure 3.86 Block diagram of a receiver

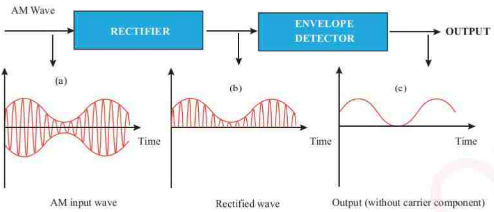


Figure 3.87 Block diagram of a detector for AM signal

Note: *The detector or demodulator also rectifies the modulated signals i.e. the negative half of modulated signal is eliminated since positive and negative modulated signals are exactly equal, the average current is zero, hence speaker cannot respond.*

The modulated signal of the form given in Figure 3.87 (a) is passed through a rectifier to produce the output as shown in Figure 3.87 (b). This envelope of signal is the message signal which is retrieved through an envelope detector (which may consist of simple R-C circuit). The output is shown in Figure 3.87 (c).
Note that the quantity on the vertical-axis can be current or voltage.

Television transmission

The television transmitter block diagram, Figure 3.88 is divided into two separate sections; First, the section that generates an electronic signal called video signal which corresponds to the actual picture. The video signal modulates a radio frequency carrier so as to be transmitted through transmitting antenna. Second, the audio signal section; this generates an electronic signal which contains sound information. The audio signal modulates another radio frequency carrier and thereafter is applied to the common transmitting antenna.

A television transmitter contains an audio modulator and video modulator. An audio signal frequency modulates another radio frequency carrier in audio modulator, while in video modulator, the video signal amplitude modulates video frequency carrier. Video signals is generated by a television camera which is displaying the scene to be televised, while the audio signal is produced by an input transducer such as microphone. However, in actual practice, there may be a number of televisions, cameras displaying the scene and a number of microphones picking up sound. The audio and video signals are amplified in audio and video amplifiers before their respective modulations. The modulated signals are joined in a combining network before they are applied to common transmitting antenna system.

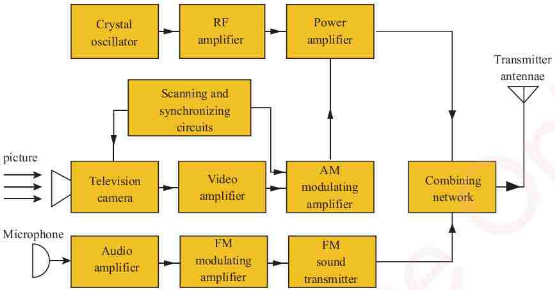


Figure 3.88 Basic monochrome television transmitter

Amplitude modulation is used for video signals since a video signal transmitter employs an AM transmitter. Frequency modulation is used for sound information since the audio signal transmitter employs FM modulator. Scanning circuits are used to make the electron beams scan the actual picture to produce the equivalent video signal. The scanning by electron beam is also found in the receiver. The beam scans the picture tube to reproduce the original picture from the video signal and this scanning at the receiver must be corresponding to the scanning at the transmitter. It is for this purpose that synchronizing circuits are used at the transmitter and at the receiver.

Television reception

The television reception block diagram as illustrated in Figure 3.89 is divided into two separate sections namely picture reception and audio reception. In picture reception section the receiving antenna captured the radiated picture and sound carrier signals from the transmitting antenna. These signals are fed to a radio frequency (RF) tuner to obtain tuned frequency. Furthermore, the signals are processed in the intermediate frequency (IF) amplifier where the carrier frequency are changed to a lower frequency. Its output is fed to a video detector or demodulator which separates video signals and audio signals from the carrier signals. The demodulated video signals are weak and hence are amplified by the video amplifier which is coupled to a picture tube to convert the electrical signal back into picture elements. Scanning circuits are used to make the electron beams scan the video signal. The beam scans the picture tube to reproduce the original picture from the video signal.

In the audio reception section, the audio signals are separated from the picture signals in the video detector stage. Also the audio carrier frequency is changed to a lower frequency in sound intermediate

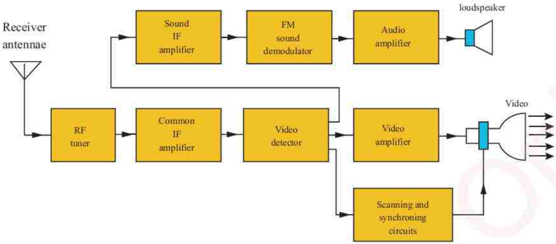


Figure 3.89 Block diagram for basic monochrome television receiver

frequency amplifier. The output from this amplifier is fed to the FM sound demodulator which separates the audio signal from the modulated signal. The demodulated audio signal is weak; it has to be amplified by the audio amplifier and then fed to the loudspeaker for reproduction into audio signals.

Activity 3.9

Visit any TV or radio station:

(i) Observe and identify the transmission and reception components

(ii) Describe the whole system of communication

Exercise 3.4

1. Define the term modulation and discuss different types of modulation.
2. Distinguish between amplitude modulation and frequency modulation.
3. Give advantages and disadvantages of AM over FM.
4. Describe the components of a communication system.
5. By using block diagram, describe the transmission and reception of TV signals.
6. Discuss the advantages of digital over analogue modes of presenting information.
7. Explain what is meant by bandwidth. Speech signals in the frequency range 300 Hz – 3400 Hz are used to amplitude-modulate a carrier wave of frequency 200 kHz. Determine:
   (a) the bandwidth of the resultant modulated signals;
   (b) the frequency range of the lower sideband; and
   (c) the frequency range of the upper side band.
8. The maximum peak - to -peak voltage of an AM wave is 16 mV while the minimum peak -to -peak is 8 mV. Find the modulation index.

Revision Exercise

1. State and explain the energy bands in solids. What are the essential features of the energy bands in solids?
2. With examples, discuss the donor and acceptor impurities. How do the impurities help in conduction of electricity?
3. Why does the electrical conductivity of an intrinsic semiconductor increase as the temperature rises?
4. Explain the concept of hole in semiconductors.Discuss the mechanism by which the hole contributes to the conduction of electricity.
5. Discuss the Zener breakdown and avalanche effect.
6. What is meant by transistor configuration? With the help of diagrams discuss the features of different transistor configurations.
7. With the help of a circuit diagram, briefly explain how p-n junction diodes can be utilized to produce full wave rectification. If the frequency of input *a.c.* signal is '*f*', what is the frequency of output signal?
8. Compare and contrast: Transistor as an amplifier and as a switch.
9. Discuss the factors that contribute to the decrease in efficiency of the transistor. Explain how to overcome those factors.
10. A transistor is used in a common emitter configuration in an amplifier circuit. When a signal of 20 m*V* is applied in the base emitter circuit, the base current changes by 20 μA and the collector current changes by 2 mA. If the load resistance is 5 kΩ find:
    (a) the current amplification;
    (b) the input resistance; and
    (c) the voltage gain.
11. Figure 3.90 shows a pnp transistor in a potential divider bias configuration. Calculate:
    (i) emitter current
    (ii) the value of common emitter voltage.

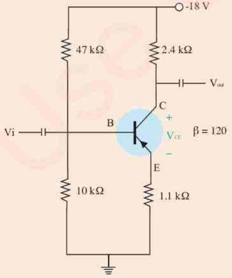


Figure 3.90 Transistor circuit

12. Name the types of transistor characteristics. Use a sketch diagram to discuss the characteristics.
13. A big saw for cutting wood at a timber store is designed to operate by switching on an electric motor only when a safety guard is lowered completely round the saw. In this case, a green light comes on when the motor is switched on. If the safety guard is not lowered, a red light comes on when the machine is switched on.

(a) Construct a truth table for a logic gate system showing the motor switched ON as HIGH (1) input and OFF as LOW (0) input; the safety guard input as HIGH (1) when lowered and LOW (0) when not lowered and the outputs HIGH (1) for switching on the lights and LOW (0) for not switching.

(b) Draw a logic gate system showing how the green and red lights can be operated as required.

14. Use De Morgan's theorem to simply the following Boolean expression.

(a) $Q=\overline{\left(X+\overline{Y}+Z\right)+\left(Y+\overline{Z}\right)}$

(b) $Q=\overline{X}+\left(\overline{Y}Z+Y\overline{Z}\right)$

15. Use the knowledge of truth tables to prove the following.

(a) $A+AB=A$

(b) $A+\overline{A}B=A+B$

(c) $\left(A+B\right)\left(A+C\right)=A+BC$

16. Deduce the Boolean expression represented by the logical gates circuit shown in Figure 3.91 and use the result to draw the simplest logic gates circuit.

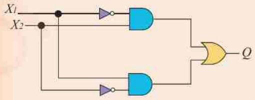


Figure 3.91 Logic gates circuit

17. What is a closed loop voltage gain as it is applied in op-amps? Explain its advantages.

18. What is the output waveform of an integrator circuit if the input waveform is a square?

19. Determine the input and output impedance of an amplifier in Figure 3.92 if Z_{in} = 2 mΩ and Z_{out} = 75 Ω, R_f=100 kΩ, R_1 = 5 KΩ and open loop gain, $A=10^5$. What will be the closed loop voltage gain?

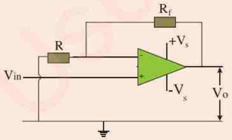


Figure 3.92 Non-inverting operational amplifier

20. (a) Explain the inverting and non-inverting operational amplifier.

(b) An operational amplifier have a voltage gain of 150. If R_1 is connected to negative terminal of operational amplifier and R_2 is a negative feedback resistance. Show that in non-inverting and inverting operational amplifier $R_1 = -2R_2$.

21. If light falls on LDR in the circuit of Figure 3.93 current flow through R_1 and causes a voltage at the non-inverting input of the operational amplifier.

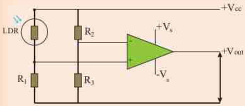


Figure 3.93 Op-amp circuit

What happens when this voltage equals the voltage set by the values of R_2 and R_3 at the inverting input? What happens to the output if R_1 and LDR are interchanged?

22. A message signal of frequency 10 kHz and peak voltage of 10 volts is used to modulate a carrier of frequency 1 MHz and peak voltage of 20 volts. Determine:
    (a) the modulation index; and
    (b) the side bands produced.
23. A transmitter radiates a total power of 10 kW. Its carrier is modulated to a depth of 60 %. Calculate;
    (a) The power of the carrier.
    (b) The power in each sideband.
24. (a) State what bandwidth is required for a radio wave of frequency 500 kHz when amplitude-modulated by audio frequencies of up to 15 kHz. Justify your answer.
    (b) What would be the effect on an audio signal which included frequencies of up to 15 kHz? Explain how the answer is obtained.
    (c) What would be the effect on an audio signal which included frequencies up to 15 kHz and transmitted by an amplitude-modulated radio signal with a bandwidth considerably less than that stated in (a)?
25. In the process of transmission of message/information signal, the signal is accompanied with noise anywhere between the information source and the receiving end. Discuss the sources of the noise.

Chapter Four

Modern physics

Introduction

Modern physics is the post Newtonian perception of physics on phenomena which cannot be explained using classical concepts. The description of nature requires theories that incorporate elements of quantum mechanics to explain physical phenomena, and provide an interpretation of experimental results. In this chapter, you will learn how quantum mechanics explains the photoelectric effect, experimental observation of hydrogen atom, and how the laser light is produced. You will also learn how quantum mechanics explain experimental observation of the hydrogen spectrum, structure of atomic nucleus, radioactivity, and the production of nuclear energy.

4.1 Quantum physics

Quantum mechanics emerged in the early 20th century from attempts to explain some properties of blackbody radiation, atomic spectra, light matter interactions and behaviour of matter on the microscopic level. It is then clear that classical physics was unable to explain these phenomena. Not only did classical predictions disagree with experiments, but even the mere existence of atoms seemed to be in the framework of classical physics. In this section some failures of classical physics including black body radiation, photoelectric effect and line spectra of the hydrogen atom and how quantum physics resolves them will be discussed.

4.1.1 Blackbody radiation

The experimental distribution of the blackbody radiation is presented in Figure 4.1 (dotted line). The continuous line is the classical attempt to describe this distribution by Rayleigh–Jeans law. In this law, radiation was treated as waves and emission of radiation was a continuous process. It is evident from this figure that, in the limit of short wavelengths, this law predicts infinite radiation intensity $I(\lambda, T)$ at a given temperature *(T)*, which is definitely inconsistent with the experimental data in which radiation intensity has finite values in the ultraviolet region of the spectrum. The discrepancy between the classical prediction and experimental data was a failure, which is commonly known as the *ultraviolet catastrophe*.

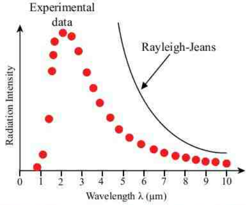


Figure 4.1 Experimental and classical prediction of blackbody radiation

As described above, the blackbody radiation curve was known experimentally, but its shape continued to elude physical explanation until when Max Planck resolved the problem in 1900 using quantum physics. The findings on blackbody radiation led Planck to postulate that radiant energy is *quantized* i.e. emitted in packets of energy called photons. He assumed that the energy of an oscillator can have only discrete energy $\left(E_n\right)$, given as:

$$E_n = nhf \qquad (4.1)$$

where $h = 6.63\times10^{-34}$ Js is the Planck's constant and f is the frequency of vibration and n is a positive integer starting from 1; this implies that the energy $\left(E_n\right)$ is quantized. In order to have the units of the horizontal axis in Figure 4.1, equation (4.1) can be expressed as:

$$E_n = \frac{nhc}{\lambda} \qquad (4.2)$$

Planck's hypothesis gives the theoretical expression for the power intensity $I\left(\lambda, T\right)$ of emitted radiation per unit wavelength as follows:

$$I\left(\lambda, T\right) = \frac{2\pi hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda k_B T} - 1} \qquad (4.3)$$

where c is the speed of light in vacuum and $k_B = 1.380\times10^{-23}$ J/K is Boltzmann's constant. The theoretical formula expressed in equation (4.3) is called *Planck's law of blackbody radiation.* This law is in total agreement with the experimental blackbody radiation (dotted curve in Figure 4.2. emitted from a surface of temperature say T.

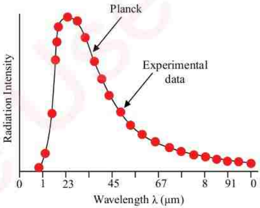


Figure 4.2 Experimental result of blackbody radiation

4.1.2 The photoelectric effect

In the photoelectric effect, electrons are emitted from matter as a result of absorption of energy from electromagnetic radiation, such as X-rays, gamma rays, visible or ultraviolet light or in some cases infrared rays. Electrons emitted in this manner are called *photoelectrons* and the phenomenon is called *photoelectric effect*. This section describes physical characteristics or observed facts of the photoelectric effect, why classical theory of light fails to explain such effects and how quantum theory can be used to resolve this failure.

Experimental study of photoelectric effect

The most common method used to study photoelectric effect is to have an evacuated tube containing a cathode (C) or emitter and anode *(A)* as a collector of electrons. The anode is connected to the positive terminal of a battery *(B)* as shown in Figure 4.3. The experiment is conducted in a dark room to avoid light falling on the cathode. As soon as the cathode *(C)* is illuminated by light of known frequency from a source *(S)*, the microammeter $(\propto A)$ registers a current.

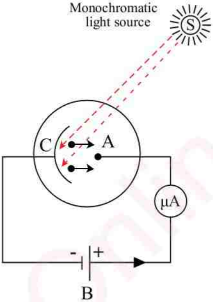


Figure 4.3 Current produced by monochromatic light in photoelectric effect

The laboratory setup of photoelectric experiment is presented in Figure 4.4, the evacuated tube has cathode and anode with a window of quartz to allow light rays to fall on the cathode. The anode and cathode are connected to voltmeter *(V)* with zero center so as to know by how much the cathode is positive or negative with respect to the anode. To provide necessary voltage, a battery *B* is connected across a potential divider *(MN)*. The cathode *(C)* is connected to the center terminal of the potential divider while the sliding contact *L* is connected to anode *(A)*. The current in the circuit is recorded by a microammeter $(\propto A)$.

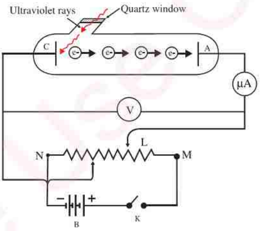


Figure 4.4 Experimental setup for photoelectric effect investigations

The experimental setup described above is then used to make the following investigation on photoelectric effect.

(i) ***Effect of light intensity on current.*** In this experiment the anode is kept at a suitable high positive potential by adjusting L along the potential divider. The intensity of light is then decreased by using filters and each time the intensity is changed the current is recorded. A plot of current versus light intensity is presented in Figure 4.5.

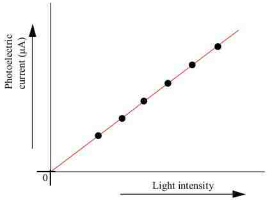


Figure 4.5 Variation of light intensity with current

(ii) Effect of potential on current. By sliding L towards N the potential on A can be reduced to zero. At this stage only electrons emitted with sufficient kinetic energy will reach A and cause the current to be registered. As L is moved towards N the current decreases to zero at potential $\left(-V_o\right)$. V_o is called the *stopping potential* because it stops energetic electrons from reaching the anode A. Figure 4.6 shows how current varies as potential is decreased from positive values (accelerating) to zero (no potential) to negative values (retarding potential) till no current flows for two light intensities. The intensity of curve Q was twice the intensity used to get curve P.

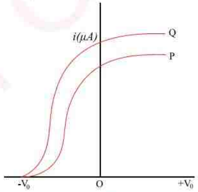


Figure 4.6 Variation of potential with current

(iii) Effect of frequency on stopping potential. Light was shone on the cathode and L was moved towards N till no current was recorded. This was repeated using different sources of light example light emitted by potassium, lithium or sodium. The frequency of light plotted against the stopping potential (V_o) is presented in Figure 4.7.

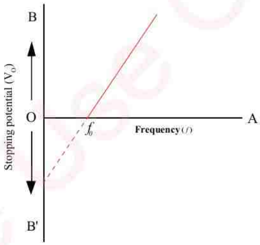


Figure 4.7 Variation of frequency with stopping potential

Laws of photoelectric emission

The experimental results obtained in photoelectric emission led to the formulation of the following laws of photoelectric emission.

(i) For a given metal there is a minimum frequency called the threshold frequency (f_o) below which no electron emission is possible no matter how intense the radiation may be or how long it takes to expose the metal to radiation.

(ii) For frequency f greater than the threshold frequency f_o, photoelectric current is directly proportional to the intensity of light.

(iii) For frequency $f > f_o$ the maximum kinetic energyis(K_{max}) directly proportional only to the frequency of incident radiation and is independent of intensity of radiation and the radiation exposure time.

(iv) The photoelectric emission is an instantaneous process since the time between exposure and emission of electron is of the order of 10^{-9} second.

Failure of wave theory to explain photoelectric effect

To get a clear picture of the failure, examination of the photoelectric effect using wave theory of light is considered. From classical physics light as a wave has two important properties i.e. its intensity and wavelength or frequency. Unfortunately when these quantities are varied, the wave theory could not explain various experimental observations of the photoelectric effect.

(i) Assuming radiation is made of waves, the greater the intensity the greater the energy of the incident radiation on the metal. The wave theory which relate intensity to energy fails to explain why velocity of emitted electrons is independent of light intensity in photoelectric experiments.

(ii) According to the classical theory of light i.e. waves, intensity of radiation is dependent on its frequency rather than amplitude of electric field vector. This perception is not in line with the observed dependency of velocity or kinetic energy of photoelectrons on frequency. Experimental results show that kinetic energy increase with frequency.

(iii) The wave theory predicts that electrons will always be emitted from a metal by radiation of any frequency if the incident beam has sufficient energy. Experiments showed that regardless of the intensity, no electrons are emitted if the frequency of radiation is below a threshold frequency.

(iv) The wave theory predicts the energy of radiation to spread continuously over the wave fronts of incident radiation. Therefore, a single electron intercept only a small portion of energy in the wave. As a results a considerable more time would be needed for the electron to absorb sufficient energy for it to be released from the surface of the metal. The experiment described above show that the electrons are emitted as soon as radiation of appropriate frequency falls on the metal. In other words, photoelectric emission is an instantaneous event that cannot be explained using classical theory of waves.

Photon theory of light

In an attempt to resolve the failure of classical theory in explaining photoelectric effect, in 1905 Albert Einstein extended the Max Planck's quantum theory by assuming radiation to be transmitted by tiny particles called *photons*. A photon in this context is perceived as a massless bundle of electromagnetic energy given by:

$$E = hf \qquad (4.4)$$

For monochromatic radiation all photons are characterized by this energy and therefore increasing intensity of the beam simply means the number of photons in the beam increases. Albert Einstein visualized emission of electrons from the metal as a result of collision with a single photon. Since all the photons energy is transferred to electrons, the photon stop to exist after energy transfer. If the energy hf of the photon is less than the energy that binds electrons to the metal, electrons will not be emitted. The binding energy is called *work function* ϕ_o of the metal. Thus the emission requirements in terms of available energy is given by:

$$E = hf = \phi_o + K_{\max} \qquad (4.5)$$

This means that the incident photons must have sufficient energy to overcome attraction between the electron and the nucleus if an electron is to be emitted. The work function can be considered as positive energy required to overcome the attraction between the electron and the nucleus, which is indicated by ϕ_1 and ϕ_2 in Figure 4.8 for two different metals.

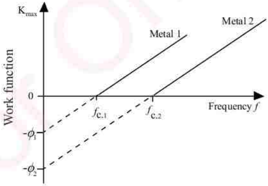


Figure 4.8 Work function and threshold frequency of two different metals

The work functions of selected different metals are displayed in Table 4.1. The remainder of the energy appears as kinetic energy $K_{\max}$ of the emitted electron as shown in equation (4.5).

Table 4.1 *Values of work functions of selected metals*

Metal	Work function ϕ *(eV)*
Na	2.46
Al	4.08
Pb	4.14
Zn	4.31
Fe	4.50
Cu	4.70
Ag	4.73
Pt	6.35

The kinetic energy $K_{\max} = \frac{1}{2}m_e v^2$ and therefore the velocity of the emitted photoelectron is given by: $V = \sqrt{\frac{2(hf - \phi_o)}{m_e}}$

When $v = 0$ the work function is given by $\phi_o = hf_0$ where f_o is the threshold frequency. Therefore:

$$f_o = \frac{\phi_o}{h} \qquad (4.6)$$

From equation (4.6) it is evident that, the threshold frequency f_o depends only on the work function ϕ_o of the metal as listed in Table 4.1 since the Planck's constant h is a *universal constant*.

Thus, photons with larger frequencies than the threshold frequency f_o are assured to produce photoelectrons because $K_{\max}$ is always greater than zero. The converse is true when the frequency of the incident photons is smaller than f_o; in this situation, the photons do not have sufficient energy to produce photoelectrons. Observing that frequency f and wavelength λ of light are related by $\lambda f = c$, where c is the speed of light in vacuum, then the threshold frequency can be expressed as:

$$\lambda_o = \frac{c}{f_o} = \frac{hc}{\phi_o} \qquad (4.7)$$

where the constant, $hc = 1240\ eVnm$. When the frequency of incident photon is small than f_o then wavelength of the incident radiation is longer than the threshold wavelength λ_o. Under this condition photoelectric effect does not occur.

If photons of energy hf strike a metal surface with work function ϕ_o, the emitted photoelectrons will have maximum kinetic energy related by $K_{max} = hf - \phi_o$. If increasing retarding potential is applied on the electrons, the velocity will gradually decrease to zero. The potential that produces this effect is called stopping potential V_o which mathematically, given as $K_{max} = eV_S$.

Important relations

(i) Einstein photoelectric equation

$$hf = \frac{hc}{\lambda} = \phi_o + \frac{1}{2}mv_{max}^2$$

(ii) If $v = 0$ then the work function ϕ is given by;

$$hf_o = \frac{hc}{\lambda_o} = \phi_o$$

(iii) The maximum kinetic energy can be obtained from;

$$K_{max} = hc\left(\frac{1}{\lambda} - \frac{1}{\lambda_o}\right)$$

$$K_{max} = h\left(f - f_o\right)$$

(iv) $V_{max} = \sqrt{\dfrac{2hc}{m_e}\left(\dfrac{1}{\lambda} - \dfrac{1}{\lambda_o}\right)}$

(v) Stopping potential V_o and frequency f

$$hf = \frac{hc}{\lambda} = \phi_o + eV_S$$

$$V_S = \frac{hf}{e} - \frac{\phi_o}{e}$$

$$V_S = \frac{hc}{e}\left(\frac{1}{\lambda} - \frac{1}{\lambda_o}\right)$$

$$V_S = \frac{h}{e}\left(f - f_o\right)$$

(vi) $= \dfrac{\left(6.63\times10^{-34}\,\text{Js}\right)\times\left(2.9979\times10^{8}\,\text{m/s}\right)}{1.602\times10^{-19}\,\text{J/eV}\times\lambda}$

$$\frac{hc}{\lambda} = \frac{1241\,\text{eVnm}}{\lambda}$$

Example 4.1

(a) The red light from helium-neon has a frequency of 4.74 x 10^{14} Hz. What is the energy of one photon?

(b) Electromagnetic radiation of wavelength $\lambda = 400$ nm falls on a surface of a metal with a work function $\phi = 1.25$ eV. What is the stopping potential of this metal?

Solution

(a) $E = hf = 6.63\times10^{-34}\,\text{Js}\times4.74\times10^{14}\,\text{Hz}$

$E = 3.14\times10^{-19}\,\text{J}$

(b) The energy relation required is:

$E = V_S e + \phi_o$ where

$$E = \frac{hc}{\lambda} = \frac{1240\,\text{nmeV}}{400\,\text{nm}} = 3.1\,\text{eV}$$

Using this value and the work function $\phi_o = 1.25\,\text{eV}$ we obtain;

$$V_S = \frac{E-\phi_o}{e} = \frac{(3.1\text{-}1.25)\text{eV}}{\text{e}}$$

$$\therefore V_S = 1.85\text{V}$$

Example 4.2

A photo-emissive surface has a threshold wavelength of 750 nm. Use this information to calculate:

(a) threshold frequency;
(b) work function in eV; and
(c) maximum speed of the electron emitted by light of wavelength 500 nm

Solution

(a) From $f_o = \dfrac{c}{\lambda_o} = \dfrac{2.9979\times10^8\text{ m/s}}{750\text{ nm}\times10^{-9}\text{m/nm}}$

$$f_o = 3.997\times10^{14}\text{ Hz}$$

(b) Work function

$$\phi = hf_o$$

$$= 6.63\times10^{-34}\text{Js}\times3.997\times10^{14}s^{-1}$$

$$= 2.65\times10^{-19}\text{J}$$

$$\phi = \frac{2.65\times10^{-19}\text{ J}}{1.602\times10^{-19}\text{ J/eV}}$$

$$\phi = 1.65\text{eV}$$

(c) Maximum speed could be obtained from the maximum kinetic energy

$$K_{\max} = \frac{1}{2}mv^2_{\max}$$

$$\frac{1}{2}mv^2_{\max} = hf - \phi = \frac{1240\text{ eVnm}}{500\text{ nm}} - 1.65\text{eV}$$

$$\frac{1}{2}mv^2_{\max} = 0.83\text{eV}\times1.602\times10^{-19}\text{ J/eV}$$

$$\frac{1}{2}mv^2_{\max} = 1.33\times10^{-19}\text{ J}$$

$$v_{\max} = \sqrt{\frac{2\times1.33\times10^{-19}\text{ kgm}^2/\text{s}^2}{9.1\times10^{-31}\text{kg}}}$$

$$v_{\max} = 5.4\times10^5\text{ m/s}$$

Example 4.3

Monochromatic light of wavelength 4500 Å is incident on a sodium surface of work function 2.3 eV. Determine:

(a) the energy of incident photons;
(b) maximum kinetic energy of the emitted electrons; and
(c) the stopping potential of sodium.

Solution

(a) $E = hf = \dfrac{hc}{\lambda}$

$$= \frac{6.63\times10^{-34}\text{Js}\times2.9979\times10^8\text{m/s}}{4500\times10^{-10}\text{m}}$$

$$E = 4.42\times10^{-19}\text{J}$$

Alternatively;

$$E = hf = \frac{hc}{\lambda} = \frac{1240\text{eVnm}}{450\text{ nm}}$$

$$E = 2.8\text{eV}$$

(b) Maximum kinetic energy,

$$K_{\max} = \frac{hc}{\lambda} - \phi = \frac{1240\text{eVnm}}{450\text{nm}} - 2.3\text{eV}$$

$$K_{\max} = 0.5\text{ eV}$$

or

$$K_{\max} = 0.5\text{eV}\times1.602\times10^{-19}\text{J/eV}$$

$$K_{\max} = 8.0\times10^{-20}\text{J}$$

Physics Form VI X.indd 181 10/11/2019 17:22

(c) Stopping potential of sodium;

$$V_s = \frac{E-\phi}{e} = \frac{(2.8\text{-}2.3)\text{eV}}{\text{e}}$$

$$V_s = 0.5\text{V}$$

Example 4.4

Monochromatic light of wavelength 300 nm is incident normally on 4 cm² surface. If the intensity of light is 15×10^{-2} W/m², find the photon flux Ψ (number of photons striking the surface per second).

Solution

The energy of each photon is given by;

$$E_{ph} = \frac{hc}{\lambda} = \frac{1240\,\text{eVnm}}{300\,\text{nm}}$$

$$E_{ph} = 4.133\text{eV/photon}$$

Total energy flowing per second (Flux) E_T

$$E_T = \left(15\times10^{-2}\ \text{J/sm}^2\right)\times\left(\frac{1}{1.602\times10^{-19}\text{J/eV}}\right)\times\left(4.0\times10^{-4}\text{m}^2\right)$$

$$E_T = 3.745\times10^{14}\ \text{eV/s}$$

The number of photon striking the surface per second also known as photon flux is given as:

$$\Psi = \frac{E_T}{E_{ph}} = \frac{3.745\times10^{14}\ \text{eV/s}}{4.133\ \text{eV/photon}}$$

$$= 9.062\times10^{13}\text{photons/s}$$

$$\Psi = 9.062\times10^{13}\text{photons/s}$$

4.1.3 De Broglie wavelength

To explain the photoelectric effect, Albert Einstein in 1905 proposed that electromagnetic waves should be considered as particles. In 1924, Louis de Broglie thought that consideration of electron and other particles of matter as waves would resolve some failures of classical physics to explain experimental observations such as the hydrogen spectrum. Later G. J. Davisson and L.H.Germer (1927) found that a beam of electrons when reflected from a metalic crystal shows diffraction patterns similar to those of electromagnetic waves. This verified the work of de Broglie that electrons behave both as waves and as particles.

According to Einstein, energy associated with wave is given by:

$$E = mc^2 \tag{4.8}$$

But energy is also written as:

$$E = hf = \frac{hc}{\lambda} \tag{4.9}$$

Electron as a particle travelling with a speed of light would have wavelength given by:

$$\frac{hc}{\lambda} = mc^2$$

$$\lambda = \frac{h}{mc} \tag{4.10}$$

Since the electron has a rest mass m it does not travel with the speed of light: instead it travels at a lower speed v, and therefore equation (4.10) becomes:

$$\lambda = \frac{h}{mv} = \frac{h}{p} \tag{4.11}$$

where p is the momentum and λ is called the de Broglie wavelength. The perception by de Broglie that matter could be a wave or a particle is frequently called *wave-particle duality,* which is the basis of modern quantum mechanics, also known as wave mechanics.

Example 4.5

Calculate the de Broglie wavelength for an electron and proton moving at $10^5\,\text{m/s}$. Given that mass of electron is $9.1\times10^{-31}\,\text{kg}$ mass of proton is $1.67\times10^{-27}\,\text{kg}$ and $h = 6.63\times10^{-34}\,\text{Js}$.

Solution

$$\lambda_e = \frac{h}{mv} = \frac{6.63\times10^{-34}\,\text{kgm}^2\text{s/s}^2}{9.1\times10^{-31}\text{kg}\times10^5\,\text{m/s}}$$

$$\lambda_e = 7.3\,\text{nm}$$

$$\lambda_p = \frac{h}{mv} = \frac{6.63\times10^{-34}\,\text{kgm}^2\text{s/s}^2}{1.67\times10^{-27}\,\text{kg}\times10^5\,\text{m/s}}$$

$$\lambda_p = 3.97\times10^{-3}\text{nm}$$

Example 4.6

(a) An electron is accelerated through a potential difference of V volts. What is the de Broglie wavelength?

(b) Determine the wavelength if $V = 100\,\text{volts}$.

Solution

(a) Work done on the electron is equal to the kinetic energy gained

$$eV = \frac{1}{2}mv^2$$

$$v = \sqrt{\frac{2eV}{m}}$$

The de Broglie wavelength for the electron is given by:

$$\lambda = \frac{h}{p} = \frac{h}{mv}$$

$$\lambda = \frac{h}{m\sqrt{\frac{2eV}{m}}} = \frac{h}{\sqrt{2meV}}$$

$$\lambda = \frac{h}{\sqrt{2meV}}$$

$$= \frac{6.63\times10^{-34}\,\text{Js}}{\sqrt{2\times1.602\times10^{-19}\text{C}\times9.1\times10^{-31}\text{kg}\times\text{V}}}$$

$$\lambda = \frac{1.228\,\text{nm}}{\sqrt{V}}$$

(b) If , V=100 volts, then:

$$\lambda = \frac{1.228\text{nm}}{\sqrt{100\,\text{V}}}$$

$$\lambda = 0.1228\,\text{nm}$$

Example 4.7

(a) Calculate the de Broglie wavelength of a bullet of mass 100 g travelling at a speed of 500 m/s and an electron travelling at same speed.

(b) How do you relate the answers to real world?

Solution

(a) $$\lambda_b = \frac{h}{p} = \frac{h}{m_b v}$$

$$= \frac{6.63\times10^{-34}\,\text{kgsm}^2/\text{s}^2}{0.1\text{kg}\times500\,\text{m/s}}$$

$$\lambda_b = 1.326\times10^{-26}\,\text{nm}$$

$$\lambda_b = \frac{h}{p} = \frac{h}{m_e v} = \frac{6.63\times10^{-34}\,\text{kgsm}^2/\text{s}^2}{9.1\times10^{-31}\,\text{kg}\times500\,\text{m/s}}$$

$$\lambda_b = 1457\,\text{nm}$$

(b) The wavelength of the large object is much smaller than small objects and hence more difficult to observe experimentally than that of smaller objects.

Example 4.8

Compute the de Broglie wavelength of an electron and neutron each travelling at 10^5 m/s

Solution

$$\lambda_e = \frac{h}{p} = \frac{h}{m_e v}$$

$$= \frac{6.63 \times 10^{-34}\,\text{kgsm}^2/\text{s}^2}{9.1 \times 10^{-31}\,\text{kg} \times 10^5\,\text{m/s}}$$

$$\lambda_e = 7.29\,\text{nm}$$

$$\lambda_n = \frac{h}{p} = \frac{h}{m_n v}$$

$$= \frac{6.63 \times 10^{-34}\,\text{kgsm}^2/\text{s}^2}{1.67 \times 10^{-27}\text{kg} \times 10^5\,\text{m/s}}$$

$$\lambda_n = 3.97\,\text{pm}$$

4.1.4 Production of X-rays

X-rays are produced when fast moving electrons strike a metal target of high atomic number like tungsten. Since electrons produce heat when they strike a target, it is necessary for the target to have a high melting point. The electrons are produced at the heated filament called cathode as shown in Figure 4.9.

There are two types of X-rays spectrum characteristics namely continuous spectrum and line spectrum characteristics.

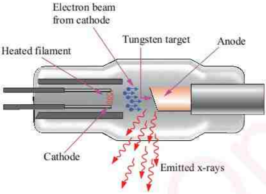


Figure 4.9 X-ray tube showing source of electrons and the target

According to the classical theory of electromagnetism, charged particles which undergo acceleration or retardation emits radiation known as *bremsstrahlung* or *braking radiation*. This radiation forms the continuous part of the X-rays shown in Figure 4.10. According to equation (4.9), the minimum wavelength λ of the continuous spectrum is readily explained by the quantum theory as:

$$\lambda_{\min} = \frac{hc}{K_{max}} = \frac{hc}{eV} \qquad (4.12)$$

where V is the accelerating potential and e is the charge of an electron. It can be shown that the accelerating potential 35 kV shown in Figure 4.10 produces $\lambda_{\min} = 0.035\,\text{nm}$. There is no classical description of the characteristic X-rays K_α and K_β superimposed on the continuous spectrum.

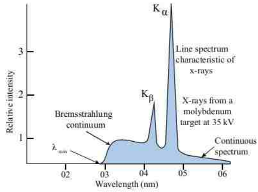


Figure 4.10 Continuous and line characteristics of x-rays spectra

The quantum description of the characteristic X-rays will now be presented. When metals of high atomic number are bombarded by a high-energy electron beam, the electrons displace electrons from deep energy levels called shells. As will be described in Bohr theory, the energy level K, L, M, N of both light and heavy atoms, are arranged as shown in Figure 4.11.

The K-series of X-ray lines are produced when the energetic electrons knock-out electrons in the K-shell. The transitions that produce K_α , K_β and K_γ characteristics X-rays are indicated by vertical lines that end in the K-energy level. If the energetic electrons knock off electrons from the L-shell, the L-series will be produced as indicated by vertical lines that end in L-energy level.

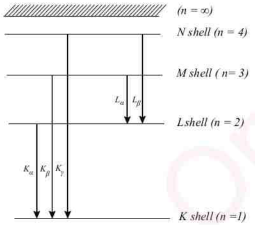


Figure 4.11 Transition of some characteristics X-rays

Uses of X-rays

X-rays have many applications, including medical diagnostics (mammography, CT scan, dental and chest X-rays), inspection of luggage at airports, inspection of the quality of welding, detection of cracks in aircraft components, and detection of the elemental composition of samples. The ability of X-rays to perform this function is depend on their passage through objects and reveal the internal structures of the objects without destruction.

Activity 4.1

Students to make a study visit to a laboratory or hospital equipped with an X-ray unit to study how an X-ray machine operates. Then students to write a comprehensive report on how an X-ray machine operate.

Example 4.9

An X-ray unit operating at 100 kV produces a current of 20 mA. Use this information to calculate:

(a) the number of electrons striking the target per second;

(b) the speed of the electron incident on the target; and

(c) the lowest wavelength of the continuous X-ray spectra.

Solution

(a) The number of electrons per unit time is related to the current as follows;

$$n = \frac{I}{e} = \frac{20\times10^{-3}\,\text{C / s}}{1.602\times10^{-19}\,\text{C}}$$

$$n = 1.25\times10^{17}\,\text{s}^{-1}$$

(b) The speed of the electron incident on the target is;

$$\frac{1}{2}mv_{\text{max}}^2 = \text{eV}_\text{S}$$

$$v_{\text{max}} = \sqrt{\frac{2eV_S}{m}} = \sqrt{\frac{2\times1.602\times10^{-19}\,C\times10^5\,\text{V}}{9.1\times10^{-31}\,\text{kg}}}$$

$$v_{\text{max}} = 1.88\times10^8\ \text{m/s}$$

(c) The lowest wavelength of the continuous X-ray spectra is given by;

$$\lambda_{\text{min}} = \frac{hc}{eV} = \frac{1240\,\text{eVnm}}{10^5\,\text{eV}}$$

$$\lambda_{\text{min}} = 0.0124\,\text{nm}$$

Exercise 4.1

1. With the help of a diagram, explain the meaning of *ultraviolet catastrophe.*
2. An electron, alpha particle, and a proton have the same kinetic energy. Which of these particles has the shortest de Broglie wavelength? Why? (Numerical calculations are not required).
3. A particle is moving three times as fast as an electron. The ratio of the de Broglie wavelength of the particle to that of an electron is 9.0778×10^{-5}. Calculate the particle's mass and identify the particle.
4. What is the value of stopping potential between the cathode and anode of a photocell, if the maximum kinetic energy of electrons emitted is 5 eV?
5. What is the largest wavelength of radiation capable of ejecting electrons from platinum whose work function is 6.3 eV?
6. Given the work functions 2 eV and 5 *eV* for metal X and Y respectively. Which metal will emit electrons when it is irradiated with light of wavelength 400 nm. Why?
7. Light from a bulb falls on a wooden table but no photon electrons are emitted. Why?
8. Visible light cannot eject photoelectrons from a copper surface whose work function is 4.4 eV. Why? Explain your answer mathematically.
9. What is the effect of increasing the intensity of light on the surface of metal?
10. Suppose a photon of wavelength 60 nm is absorbed by a hydrogen atom whose ionization energy is 13.6 eV. Find the kinetic energy of the ejected electron.

4.2 Atomic physics

The atomic theory dates back to 1803 when Dalton proposed a modern theory of the atom based on five assumptions. First, matter is made up of atoms that are indivisible and indestructible. Second, all atoms of an element are identical. Third, atoms of different elements have different weights and different chemical properties. Fourth, atoms of different elements combine in simple whole numbers to form compounds. Fifth, atoms cannot be created or destroyed. All of these assumptions still hold true today except the first one which was proved wrong by the discovery of the electron by J. J. Thomson in 1897. During this period the proton was known to have positive charge, therefore in attempt to produce neutral atom by rearranging positive protons and negative electrons J. J. Thomson proposed the *plum pudding* model of the atom in 1904. In this section it will be learnt as to why the plum pudding model proposed by J.J. Thomson was replaced by the atomic planetary model proposed by Ernest Rutherford in 1913. Also it will be understood how the atomic planetary model was used in the same year by Niels Bohr to formulate another model that can be used to describe and analyse energy levels of the hydrogen atom.

4.2.1 Rutherford planetary model

In the Rutherford scattering experiment, energetic alpha particles (about 5 MeV) were released by a radioactive source placed in a lead container with a hole in one side to produce a beam of alpha particles as indicated in Figure 4.12. The beam was directed on a thin gold foil of about $0.4\,\mu\text{m}$ diameter which scattered the incident alpha particles at different angles. The scattered alpha particles were detected by a round type phosphor screen detector. Note that alpha particles are doubly ionized helium atom. The foil was made from gold which is a most expensive metal because of its high malleability and ductility which make it possible to produce thinner foils by hammering or rolling than any other metal.

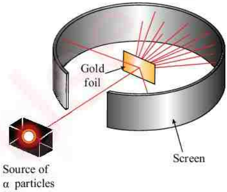


Figure 4.12 Experimental setup of Rutherford scattering experiment

Arrangement of atoms in the gold thin foil used in the Rutherford experiment is shown in Figure 4.13. If the thickness of the foil in this figure is $0.4\,\mu\text{m}$ and diameter of the gold atom is $0.0003\,\mu\text{m}$, then this thickness will consist of about 1000 gold atoms. The observations from the Rutherford experiment indicated in Figure 4.13 show that most of the alpha particles pass through the foil un-deflected, some were scattered appreciably $\left(0 < \theta \leq 90^o\right)$ and few (1 out 8000) were suffered large angle deflections $\left(90^o < \theta \leq 180^0\right)$. The observation that most of the alpha particles pass un-deflected through the foil was interpreted as the atom being mostly empty. However, the large angle deflection was surprising because it falsifies the plum pudding model of the atom by J.J. Thomson which predicted that all alpha particles would scatter at small angles. The large angle deflection suggests that all the positive charge and nearly all the mass of the atom was concentrated in a very small volume at the center of the atom which was called the nucleus. Furthermore, sufficiently negative charged electrons were revolving around the nucleus to make the atom neutral as expected. The arrangement of the electrons orbiting around the nucleus makes what is called the *planetary model or nuclear model of the atom.*

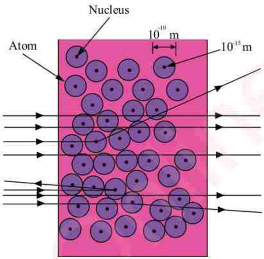


Figure 4.13 Rutherford scattering experiment

Limitations of the planetary model

Although the Rutherford atomic model was based on experimental observations it failed to explain certain fundamental observations. The proposal that the electrons revolve around the nucleus in fixed paths called orbits was not in agreement with established Maxwell's theory which stated that accelerated charged particles emit electromagnetic radiations and therefore could not explain the stability of an atom. Calculations have shown that the electron in this model would lose its energy and collapse in the nucleus in less than 10^{-8} seconds. Therefore, Rutherford theory is incomplete because it does not explain why this does not take place i.e. stability of the atom. Second limitation is that, the predicted mass of the atom by the model was consistently half the measured mass obtained from experiments. When confronted with this limitation in 1926, Rutherford resolved the limitation by proposing the neutron hypothesis. In this hypothesis Rutherford proposed the existence of a neutral particle in the nucleus with mass close to that of a proton. This hypothesis was verified experimentally by Chadwick in 1932. Third drawback is that the Rutherford model did not say anything about the arrangement of electrons in the atom. Fourth limitation was failure of the model to explain the origin of the observed spectral lines emitted by the hydrogen atom and the Rydberg R constant in equation (4.28). The subsequent atomic models were formulated to address these limitations.

4.2.2 Bohr atomic model of hydrogen atom

The Bohr model was formulated using the planetary model proposed by Rutherford while keeping in mind the inherent limitations. In this section, discussion on how some of the limitations were overcome by Bohr postulates during formulation of atomic model and the use of this model to analyse the energy levels of the hydrogen atom is made.

Bohr postulates

In formulating the atomic model of the hydrogen atom, Bohr made three postulates:

(a) Electrons revolve round the nucleus with definite velocities in concentric circular orbits situated at definite distances from the nucleus. The energy of an electron in a certain orbit remains constant. As long as it remains in that orbit, it neither emits nor absorbs energy. These are termed stationary states or main energy states.

(b) Bohr proposed that the angular momentum of an electron is quantized. Thus, the motion of an electron is restricted to those orbits where its angular momentum is an integral multiple of $\frac{h}{2\pi}$, where h is Planck's constant.

(c) The energy of an electron changes only when it moves from one energy level to another. An electronic transition from an inner orbit to outer orbit involves absorption of energy. Similarly, when an electron jumps from an outer orbit to an inner orbit it releases energy, in the form of radiation equals to the difference between the two energy levels.

Mathematical formulation of the Bohr model

The total energy E of an electron in an orbit shown in Figure 4.14 is the sum of the kinetic energy KE and potential energy PE given by:

$$E = KE + PE \qquad (4.13)$$

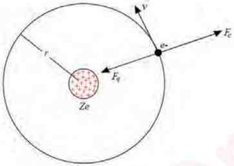


Figure 4.14 An electron performing circular orbit of radius r around nucleus of charge Ze

According to Figure 4.14 an electron of mass m_e undergoing a circular orbit of radius r with tangential velocity v has kinetic energy:

$$KE = \frac{1}{2}mv^2 \qquad (4.14)$$

Considering a nucleus as a point charge the potential at point distance r from the centre of the nucleus is $V = \frac{-e}{4\pi\varepsilon_o r}$ Thus, the work done to bring an electron from infinity becomes $V = \frac{-e^2}{4\pi\varepsilon_o r}$, where negative sign means the nucleus attracts the electron, also the potential at infinity is taken as zero. For an atom with atomic number Z the potential energy (PE) becomes:

$$PE = -\frac{(Ze)(e)}{4\pi\varepsilon_o r} \qquad (4.15)$$

The electron is held in a circular orbit of radius r under the influence of Coulomb's electrostatic force balanced by centrifugal force as follows:

$$\frac{mv^2}{r} = \frac{Ze^2}{4\pi\varepsilon_o r^2} \qquad (4.16)$$

Multiplying both sides of equation (4.16) by mr^3 we get:

$$(mvr)^2 = \frac{Ze^2 rm}{4\pi\varepsilon_o} \qquad (4.17)$$

Using the second postulate $mvr = \frac{nh}{2\pi}$ in equation (4.17) we obtain:

$$\left(\frac{nh}{2\pi}\right)^2 = \frac{Ze^2 rm}{4\pi\varepsilon_o} \qquad (4.18)$$

From equation (4.18) the radius of the nth orbit is then given by:

$$r_n = \frac{n^2 h^2 \varepsilon_o}{Z\pi me^2} \qquad (4.19)$$

The radius of the first orbit $n = 1$, for the hydrogen atom, often denoted by r_o or a_o, is called the Bohr radius, and is given by:

$$r_o = \frac{h^2\varepsilon_o}{\pi me^2} = 5.29 \times 10^{-11} \text{ m} \qquad (4.20)$$

It is interesting to note that the first orbit of singly ionized helium, the r_o can be evaluated from equation (4.19) by replacing $Z = 2$ to get;

$$r_o = \frac{h^2\varepsilon_o}{2\pi me^2} = 2.65 \times 10^{-11} \text{ m} \qquad (4.21)$$

Therefore, increase in Z reduces the radius of the first orbit.

From equation (4.16) it is clear that v needed to calculate the KE is given as;

$$v^2 = \frac{Ze^2}{4\pi m\varepsilon_o r} \qquad (4.22)$$

Using equation (4.22) in (4.14) to get KE and then equation (4.13) can be written as:

$$E = \frac{1}{2}\frac{Ze^2}{4\pi\varepsilon_o r} - \frac{Ze^2}{4\pi\varepsilon_o r} = -\frac{Ze^2}{8\pi\varepsilon_o r} \qquad (4.23)$$

Eliminating r from equation (4.23) using equation (4.19) we get;

$$E_n = -\left(\frac{me^4}{8\varepsilon_o^2 h^2}\right)\left(\frac{Z^2}{n^2}\right); \ (n = 1, 2, 3, ...) \qquad (4.24)$$

Since the energy of the electron in the nth orbit is negative, then work has to be done to free the electron. This can be achieved by adding positive energy often called ionization energy to make E_n in equation (4.24) equal to zero. This result can also be achieved when n in equation (4.24) tend to infinity. Since the terms in the first bracket are constant then:

$$E_n = -\frac{13.6 Z^2 \ eV}{n^2} \qquad (4.25)$$

Equation (4.25) as the Bohr model of hydrogen atom for $Z = 1$ has to explain experimental observations that will now be described below.

4.2.3 Success of the Bohr theory of the hydrogen atom

Energy levels

Energy levels of an atom are usually represented by horizontal lines as indicated in Figure 4.15. When the electron in the hydrogen has an energy of -13.6 eV, the atom is said to be at ground state and when it is -3.39 eV it is in the first excited state. The energy levels presented in Figure 4.15 are obtained by increasing n in equation (4.25) from $n = 1$ to $n \approx \infty$. Observe that the presence of n^2 in the denominator of equation (4.25) is responsible for unequal spacing of energy levels and, as expected, the energy of the electron in the hydrogen atom is zero when n tend to infinity.

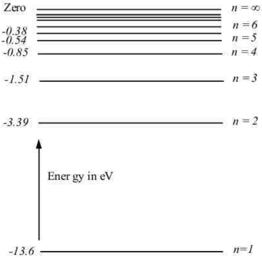


Figure 4.15 Energy levels of the hydrogen atom

Prediction of Rydberg constant

According to the 3rd postulate, the energy of the radiation emitted by the hydrogen atom when an electron jumps from an outer orbit m to an inner orbit n is given by;

$$E_m - E_n = hf \text{ where}$$

$$E_m = -\frac{13.6Z^2 \text{ eV}}{m^2} \text{ and}$$

$$E_n = -\frac{13.6Z^2 \text{ eV}}{n^2}$$

$$hf = 13.6\ eV\left(\frac{1}{n^2} - \frac{1}{m^2}\right) = \frac{hc}{\lambda} \text{ where } m > n \qquad (4.26)$$

From equation (4.26) the equivalent of equation (4.8) obtained empirically before, can now be given by;

$$\frac{1}{\lambda} = \frac{13.6\,eV}{hc}\left(\frac{1}{n^2} - \frac{1}{m^2}\right) = R_H\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \qquad (4.27)$$

Therefore, the Rydberg constant R_H predicted by the Bohr theory is:

$$R_H = \frac{13.6\text{eV}}{\text{hc}} \qquad (4.28)$$

$$R_H = \frac{13.6\text{eV}}{1240 \text{ eV.nm} \times 10^{-9} \text{m/nm}}$$

$$= 1.097677 \times 10^7 \text{ m}^{-1}$$

The Rydberg constant obtained using Bohr theory is in good agreement with the measured value of $1.097373 \times 10^7 \text{ m}^{-1}$.

Prediction of the visible hydrogen spectrum

The model is successful in predicting the visible spectrum of the hydrogen atom known as the Balmer series.

The comparison of the predicted and experimental wavelengths observed in the visible Balmer series are summarized in Table 4.2. The predicted values were obtained using the following equation;

$$\frac{1}{\lambda} = \frac{13.6\,eV}{1240\,eV\ nm}\left(\frac{1}{2^2} - \frac{1}{m^2}\right);\ m = 3,4,5,... \qquad (4.29)$$

Table 4.2 Measured and predicted wavelength of the Balmer series

m	λ-Measured	λ-Predicted	λ-Deviation
3	656.29	656.47	-0.03
4	486.13	486.27	-0.03
5	434.08	434.17	-0.02

Complete Energy spectrum of the hydrogen atom

It is important to observe that the Balmer series as described by equation (4.29) was formed by a transition from higher energy levels ending to $n = 2$. Some scientists started to question on existence of other series that start from higher energy levels m and ending to $n = 1, 3, 4, 5, ...$ as shown Figure 4.16. The series that end to $n = 1, 3, 4, 5, 6, ...$ are called Lyman, Paschen, Brackett, Pfund and Humphrey series, respectively.

Observe that, Humphrey series for n = 6 is not shown in Figure 4.16. From this sequence it is evident that there would be other series beyond the Humphrey series, however such series have not been discovered because the energies emitted are too low that existing instruments cannot detect them.

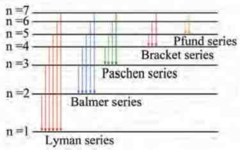


Figure 4.16: Spectral series of the hydrogen atom

4.2.4 Limitations of the Bohr theory

The Bohr theory had several limitations which will be briefly discussed as follows:

(a) The theory explains the spectra of simple atoms only e.g. the hydrogen atom and hydrogen-like atoms such as the singly ionized helium atom, and sodium with ten electrons removed. Therefore, the theory cannot be used to calculate spectral lines (similar to those shown in Figure 4.16) for atoms with many electrons.

(b) From experiment, some lines formed by the hydrogen atom have higher intensity than others. The theory cannot explain why intensities of spectral lines emitted by the hydrogen atom differ.

(c) The theory does not give information about the distribution of electrons in multi-electron atoms.

(d) The theory does not give an explanation about the electron orbit stability in the hydrogen atom.

Regarding the fourth (d) limitation above, the introduction of the concept of particle-wave duality of the electron in 1925 by de Broglie resolved the ambiguity on which the hypothesis was based. The mathematical description for a permitted stationary orbit was given by de Broglie as:

$$mvr = \frac{nh}{2\pi} \qquad (i)$$

where n is an integral called principal quantum number which makes angular momentum quantized. According to de Broglie, allowed orbits must have circumference equal to the interger multiples of the de Broglie wavelength. That means:

$$2\pi r = n\lambda = \frac{nh}{mv} \qquad (ii)$$

Re-arranging terms we get:

$$mvr = \frac{nh}{2\pi} \qquad (iii)$$

Equation (iii) is the same condition given by the Bohr quantum postulates of stable orbits. Thus, when an electron is considered as a wave, stability of electron orbits becomes easy to explain.

Example 4.10

(a) Find the total energy of the electron in the hydrogen atom in the third excited state.

(b) Show that $hc = 1240$ eVnm given that $h = 6.63 \times 10^{-34}$ Js, $e = 1.6 \times 10^{-19}\ C$ and $c = 2.9979 \times 10^{8}$ m/s

(c) A hydrogen atom is in a state whose binding energy is 0.65 eV. If the electron makes a transition to a state with an excitation energy of 10 eV, what would be the wavelength of the emitted photon?

Solution

(a) The solution require the use of equation;

$$E_n = -\frac{13.6Z^2 \text{ eV}}{\text{n}^2}$$

For third excited state $n = 4$ and therefore,

$$E_4 = -\frac{13.6 \text{ eV}}{4^2}$$

$$E_4 = 0.85\text{eV}$$

(b) $hc = 6.63\times10^{-34} \text{ Js}\times2.9979\times10^8 \ \frac{\text{m}}{\text{s}}$

$$= 1.988\times10^{-25} \text{ Jm}$$

$$hc = 1.9876\times10^{-25} \text{ Jm}\times\left(\frac{1}{10^{-9}\text{m/nm}}\right)$$

$$\times\frac{1}{1.602\times10^{-19}\text{J/ev}}$$

$$hc = 1240.70 \text{ eVnm}$$

$$hc \approx 1240\text{eVnm}$$

(c) The energy level of electron state with binding energy $0.65\,\text{eV}$ will be in energy state given by $E_2 = -0.65$ eV. By definition excitation energy is the energy required to raise an electron from the ground state to any of the excited states and is given by;

$$E_1 = \text{-13.6eV+10.0eV= -3.6eV}$$

But $\Delta E = \frac{hc}{\lambda} = E_2 - E_1$:

$$= -0.65eV - (-3.6eV) = 2.95\text{eV}$$

$$\lambda = \frac{1240\text{eVnm}}{2.95\text{eV}}$$

Therefore, $\lambda = 420.34\,\text{nm}$

Example 4.11

(a) The longest wavelength in the Lyman series for the hydrogen is 121.5 nm. Use this information to calculate the Rydberg constant.

(b) Use the Bohr theory to estimate the Rydberg constant.

Solution

(a) By definition, the Rydberg constant is given by the equation;

$$\frac{1}{\lambda} = R_H\left(\frac{1}{n^2} - \frac{1}{m^2}\right) = R_H\left(\frac{m^2 - n^2}{n^2m^2}\right)$$

where $n = 1$ and $m = 2$

Re-arranging term we get;

$$\frac{1}{R_H} = \lambda\left(\frac{m^2 - n^2}{n^2m^2}\right)$$

$$R_H = \frac{1}{\lambda}\left(\frac{n^2m^2}{m^2 - n^2}\right)$$

$$R_H = \frac{1}{121.5\times10^{-9}\,\text{m}}\left(\frac{4}{3}\right)$$

$$R_H = 1.0974\times10^7\,\text{m}^{-1}$$

(b) By definition

$$E = \frac{hc}{\lambda} = 13.6\text{eV}\left(\frac{1}{n^2} - \frac{1}{m^2}\right)$$

$$\frac{1}{\lambda} = \frac{13.6\ eV}{hc}\left(\frac{1}{n^2} - \frac{1}{m^2}\right)$$

$$R_H = \frac{13.6\,\text{eV}}{\text{hc}} = \frac{13.6\,\text{eV}}{1240\,\text{eVnm} \times 10^{-9}\,\text{m/nm}}$$

$$R_H = 1.0968 \times 10^7\,\text{m}^{-1}$$

Example 4.12

The radius of the nucleus is estimated by $R = R_o A^{1/3}$. Use this information to determine the density of ${}^{16}_{8}O$ and ${}^{208}_{82}Pb$. Comment on the magnitude and similarity of the two nuclear densities.

Solution

Volume of a nucleus is $V_N = \frac{4}{3}\pi\left(R_o A^{1/3}\right)^3$

$$V_N = \frac{4}{3}\pi R_o^3 A$$

Mass of a nucleus of mass number A is given as;

$$A(u) \times 1.67 \times 10^{-27}\,\text{kg}$$

Nuclear density is then given by the usual definition;

$$\rho_N = \frac{M}{V} = \frac{A(u) \times 1.67 \times 10^{-27}\,\text{kg}}{\frac{4}{3}\pi R_o^3 A(u)}$$

$$\rho_N = \frac{M}{V} = \frac{1.67 \times 10^{-27}\,\text{kg}}{\frac{4}{3}\pi \times \left(1.4 \times 10^{-15}\,\text{m}\right)^3}$$

$$= 1.45 \times 10^{17}\,\text{kg/m}^3$$

Since ρ_N is independent of A, nuclear density of oxygen and lead are the same and equal to $\rho_N = 1.45 \times 10^{17}\,\text{kg/m}^3$. Density of metals is of the order of $10^4\,\text{kg/m}^3$. The bulk density is much smaller than nuclear density because from the Rutherford experiment it was concluded that most of the foil was empty and almost all the mass and charge were concentrated in the nucleus, hence the enormous density of the nucleus.

Example 4.13

In a Rutherford experiment an alpha particle of kinetic energy 8 MeV makes a head on collision with a nucleus of sodium $\left({}^{23}_{11}Na\right)$. Compare the radius obtained from the collision experiment to that obtained using $R = R_o A^{1/3}$. Comment on the difference or similarity of the values obtained.

Solution

Since both alpha *(2e)* and sodium nucleus *(11e)* have positive charges, the kinetic energy will be converted to potential energy *(PE)* at closet distance *(d)*.

$$KE = PE = \frac{(2e)(11e)}{4\pi\varepsilon_o d} = 8\,\text{MeV}$$

$$d = \left(\frac{e^2}{4\pi\varepsilon_o}\right)(22)\frac{1}{(8\,\text{MeV})}$$

but

$$\frac{e^2}{4\pi\varepsilon_o} = 1.44\,\text{MeVfm}$$

$$d = (1.44\,\text{MeVfm})(22)\frac{1}{(8\,\text{MeV})}$$

$$= 3.960\,\text{fm}$$

From formula of nuclear radius

$$R = (1.44\,\text{fm}) \times (23)^{1/3} = 4.095\,\text{fm}$$

$$= 4.095\,\text{fm, where } f = 10^{-15}$$

The difference of 3.3% is small. However, it is important to note that $R = R_o A^{1/3}$ is an empirical formula derived from alpha scattering experiment.

Exercise 4.2

1. The hydrogen spectrum consists of a large number of spectral lines and yet this atom has only one electron. Why?
2. How did de Broglie use the wave particle duality of the electron to explain the basis for the existence of stable orbits in the Bohr's theory?
3. A triply ionized beryllium $\left(Be^{+++}\right)$ has the same energy as the energy of the ground state of the hydrogen atom.
   (a) What is the value of n for this state?
   (b) If an electron obtained in 3(a) makes transition to the ground state, what is the wavelength of the emitted photon?
4. Calculate the wavelength in nm of the first $\left(L_\alpha\right)$ line and second $\left(L_\beta\right)$ members of the Balmer series of the hydrogen atom.
5. Use Bohr's theory to estimate the ionization energy of the doubly ionized Lithium ion and the minimum wavelength a photon must have to cause the third ionization.

4.3 Laser

In 1917 Albert Einstein laid the foundation for laser devices when he proposed the concept of "Stimulated Emission," The concept is fundamental to the operation of all lasers. In 1950 Charles Townes, Nikolay Basov and Alexander Prokhorov developed the quantum theory of the stimulated emission and demonstrated it in microwaves.

In 1959 Gordon Gould proposed that the stimulated emission can be used to amplify light. Further, he described an optical resonator that can create a narrow beam of light and called it laser. Laser is an abbreviation for **l**ight **a**mplification by **s**timulated **e**mission of **r**adiation. Laser as a device was first built in 1960 by Theodore H. Maiman and it has a wide range of practical applications. In this section production, properties, types and applications of laser light will be discussed.

4.3.1 Production of laser

As already described in the previous section, electrons of atoms at ground state of E_1 can absorb photons of energy hf and move into the excited state E_3 as indicated in Figure 4.17. Atoms in the state E_3 which has a life time of about 10^{-8} s decay spontaneously to state E_2 which is a metastable state i.e. with longer life time of 10^{-3} s. This means that de-excitation of atoms which end in state E_2 creates a *population inversion* of electrons at level E_2. By definition, population inversion in matter is a situation in which more electrons are in high energy E_2 state than in lower energy state E_1.

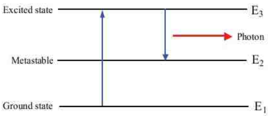


Figure 4.17 Creation of population inversion in E_2

If a stimulating photon of energy $hf = E_2 - E_1$ is introduced to the excited atom with the electrons in the metastable state the photon stimulates an electron to drop to the lower energy level E_1 thereby emitting an additional photon of energy $E_2 - E_1$. The two identical photons proceed to induce four photons and the process is repeated to produce eight photons as indicated in Figure 4.18. The eight photons encounter more electrons in the metastable state to produce sixteen photons and the process is repeated to produce a cascade or chain reaction called *stimulated emission*. Production of laser light by this process is coherent because, as observed in Figure 4.18 all photons in the stimulated emission have the same frequency. In this case laser are in phase and confined as an instrument, produces a very intense beam which is known as laser light. To summarize, the following are basic requirements to produce laser light; the existence of a metastable state in the atoms, the creation of population inversion and the stimulating and emitted photons from the metastable state must be coherent (i.e. the same frequency).

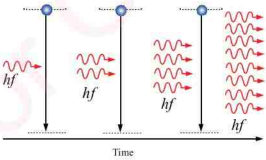


Figure 4.18 Production of laser beam

4.3.2 Properties of laser light

There are three distinct characteristics that distinguish laser light from ordinary light. For the laser beam to be intense it has to be collimated, monochromatic and coherent. A laser beam is *collimated* if it consists of waves traveling parallel to each other in a single direction with very little divergence. This property allows the laser light to be focused to a beam of very high intensity. The term *monochromatic* means that laser beam has a single colour thus single frequency, this is necessary condition to create an intense laser beam because of constructive interference. The light from a laser is *coherent,* that is, the waves of the laser light are in phase in space and time; this is a necessary condition for constructive interference.

4.3.3 Types of laser

Laser is often classified by the type of materials used for the production of laser light. The most common materials used are solid-state, gas, liquid and semiconductor.

In solid-state laser, glass or crystalline materials are used by introducing impurities in hosting material, using a process called doping. In this case, the rare earth elements such as cerium (Ce), erbium (Eu) and terbium (Tb) are most commonly used as dopants. Doping is described in detail in Chapter Three.

A gas laser has electric current discharged through a gas inside the laser medium to produce laser light. This type of lasers are commonly used in applications that require laser light with very high beam quality and coherence for long distance. In gas laser, the laser medium is made up of the mixture of gases and works on the principle of converting electrical energy into light energy. This mixture of gases packed into a glass tube acts as an active medium or laser medium. This type of laser produces a laser light of wavelength in the infrared region of around 1.15 μm.

In liquid laser light supplies energy to the laser medium. A dye laser is an example of the liquid laser. A dye as a laser uses organic liquid solution as laser medium. This device generates laser light from the excited energy states of organic liquid solvents. It produces laser light beam in the near ultraviolet and infrared regions of the spectrum.

Semiconductor lasers, also known as laser diodes, are different from solid-state lasers because in solid state lasers, light energy is used as the pump source while in semiconductor lasers, electrical energy is used as the pump source. In this type of lasers, a p-n junction of a semiconductor diode forms the active laser medium in which optical gain is produced.

Activity 4.2

The student to demonstrate using laser pointer that a laser light is collimated. This will be done by moving a screen away from the source while observing how the beam will spread with increase in distance from the source.

Activity 4.3

Monochromatic nature of laser light will be demonstrated by taking a red laser and investigate its transmission through different filters including a red filter.

4.3.4 Applications of Laser

Laser is used in various fields, including medicine, research, military, industry, and commercial fields. Some of the applications are briefly explained below.

(a) Precision cutting, welding and drilling of tiny holes into hardest materials are more easily achievable by laser beam than any other techniques.

(b) In surgery, laser has proved to be more reliable and accurate instrument for removing small tumors and making delicate operations including the eye without loss of blood or tissue damage. Laser can also be used to fragment stones in the gallbladder and kidney without harming these organs.

(c) In the film industry, laser have an exciting application in the production of true three-dimensional images called *hologram*. This is a special type of photograph or image made with a laser in which the object shown looks solid, as if it were real rather than flat. This is possible because the hologram records not only the intensity of the light but also the phase difference between two image-beams that cause the interference pattern to reconstruct an image in air that look like a real object.

(d) In the military, lasers are used in target acquisition and to destroy heavy machines by focusing the beam into the target.

(e) In the printing industry, laser beam focused and scanned across a photoactive selenium coated drum where it produces a charged pattern which mirrors the materials to be printed.

(f) In CD and optical discs analogue sound and picture signals are digitized by sampling and coded as binary numbers and stored in grooves on compact discs and DVD respectively. As the focused laser beam sweeps over the grooves, it reproduces the binary numbers which are then converted by special circuits to analogue signals which are replica of the original sound or pictures.

(g) Barcode scanners used in supermarkets are based on helium–neon laser to scan the barcodes to identify different products. The laser beam bounces off the code sending the modulated beam to a light detector and then to a computer which has information of the stored products.

Exercise 4.3

1. Explain the meaning of metastable state.
2. How is it possible to create laser beam strong enough to use as a weapon to destroy heavy machinery?
3. How does laser light differ from normal light?
4. How do satellites use laser to communicate with each other?
5. State differences between surgical laser and communication laser.

4.4 Nuclear physics

After the formulation of the neutron hypothesis in 1926 by Rutherford and after experimental confirmation of its existence in the atomic nucleus by Chadwick in 1932, the atom was considered to consist of electrons orbiting the nucleus. In the atomic physics section, it was discussed about the arrangement of the electrons around the nucleus. In this section the coverage will be on the structure of the nucleus, the relationship between nuclear mass and its binding energy, the criteria for stability and instability (radioactivity) of the nucleus, uses and hazards of radioisotopes, difference between fission and fusion and how they relate to nuclear energy production.

4.4.1 Nuclear structure

Composition of the nucleus

The nucleus of an atom consists of protons and neutrons which are collectively called nucleons because they are found in the nucleus. Figure 4.19 shows the arrangement of neutrons and protons in the nucleus. The number of protons (*atomic number*) in the nucleus is often denoted by Z and the number of neutrons in the nucleus is denoted by N. The total number of nucleons (*mass number*) is denoted by A.

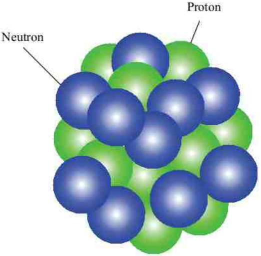


Figure 4.19 Composition of the nucleus

The atomic and mass numbers are related by:

$$A = Z + N \quad (4.30)$$

Symbolically the nucleus is represented by:

$$_{Z}^{A}X \quad (4.31)$$

where X represents the chemical element e.g. carbon (C), oxygen (O) and sodium (Na). When atoms have the same atomic number Z, such atoms are called isotopes. For example, $_{8}^{15}O, _{8}^{16}O, _{8}^{17}O$ and $_{8}^{18}O$ are isotopes of oxygen. If the atoms have the same mass number A, they are called isobars and if they have the same number of neutrons, they are called isotones. Examples of isobars are $_{11}^{23}Na$ and $_{12}^{23}Mg$ while $_{17}^{37}Cl$ and $_{19}^{39}K$ are examples of isotones.

4.4.2 Relationship between nuclear mass and binding energy

According to the mass energy equation of Einstein, a relation exists between mass and energy. In this section we will learn on how nuclear mass and binding energy are related and how this relationship influences nuclear stability.

Mass defect

The nuclei of all atoms, except hydrogen are made of combination of neutrons and protons. According to nuclear particle experiments, the measured mass *M(A,Z)* of a nucleus of mass number *A* and atomic number *Z* is always *less* than the sum of the masses of its constituent nucleons (protons and neutrons) i.e. $M(A,Z) < M = ZM_p + (A-Z)M_N$. The mass difference is commonly called **mass defect** Δm it is given by:

$$\Delta m = ZM_p + (A-Z)M_N - M(A,Z) \quad (4.32)$$

Let us use helium $_{2}^{4}\mathrm{H}_e$ to illustrate this point.

$$M(A,\ Z) = \mathrm{H_e}(4,2) = 4.0015\,\mathrm{u}$$
$$2M_p = 2(1.00728\,\mathrm{u}) = 2.01456\,\mathrm{u}$$
$$2M_N = 2(1.00866\,\mathrm{u}) = 2.01732\,\mathrm{u}$$

Using this information in equation (4.32) we get:

$$\Delta m = 2.01456\,\mathrm{u} + 2.01732\,\mathrm{u} - 4.00151\,\mathrm{u}$$
$$\Delta m = 4.03188\,\mathrm{u} - 4.00151\,\mathrm{u}$$
$$\Delta m = 0.03037\,\mathrm{u} \quad (4.33)$$

Note that the conversion factor from atomic mass unit to kilogram is $1u = 1.6605 \times 10^{-27}$kg. Therefore, splitting a nucleus of a helium atom into its individual protons and neutron a total mass will increase. This increase is what holds the nucleus together.

Binding energy

The nucleons are held together in the nucleus of an atom by the energy provided by the mass defect. This mass defect is related to the binding energy of the nucleus through Einstein mass energy relation. By definition, the binding energy is the energy required to liberate all the nucleons from the nucleus of an atom as indicated in Figure 4.20.

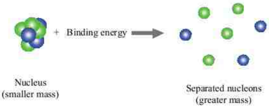


Figure 4.20 Binding energy as the energy required to break up nucleus

This binding energy *BE* is calculated by:

$$BE = \Delta mc^2 \qquad (4.34)$$

According to equation (4.32), the binding energy of a nucleus of an atom ${}^{A}_{Z}X_N$, is given by;

$$BE = \left[ZM_p + \left(A-Z\right)M_N - M\left(A,Z\right)\right]c^2 \qquad (4.35)$$

According to equations (4.34) and (4.35) the binding energy for the helium atom is

$$BE = \left(0.03037\,\text{u} \times 1.6605\times10^{-27}\,\text{kg/u}\right)\text{c}^2$$

$$BE = 5.0429\times10^{-29}\,\text{kg}\times\left(2.9979\times10^{8}\,\text{m/s}\right)^2$$

$$BE = 4.5323\times10^{-12}\,\text{J} \qquad (4.36)$$

Energy equivalent of *1u*

According to Einstein mass energy relation gives the conversion factor;

$$1\text{u}=1\text{u}\times1.6605\times10^{-27}\text{kg/u}\times\left(2.9979\times10^{8}\text{m/s}\right)^2$$

$$1u = 1.4924\times10^{-10}\,\text{J} \qquad (4.37)$$

Dividing equation (4.37) by $1.6022\times10^{-19}\,\text{J}/\text{eV}$ gives another conversion factor:

$$1\text{u}=931.5\text{MeV} \qquad (4.38)$$

Therefore, the *BE* of helium atom in *MeV* is given as;

$$BE = \frac{4.54\times10^{-12}\text{J}}{1.6022\times10^{-13}\text{J/MeV}}$$

$$BE = 28.29\ \text{MeV}$$

Alternatively, the *BE* can be obtained directly from the mass defect given by equation (4.33) as follows;

$$BE = 0.03037\,\text{u}\times\left(931.5\text{MeV/u}\right)$$

$$BE = 28.29\text{MeV}$$

From these calculations, the energy needed to separate a nucleus into individual protons and neutron to produce mass defect is called the binding energy. Binding energy is linked to size

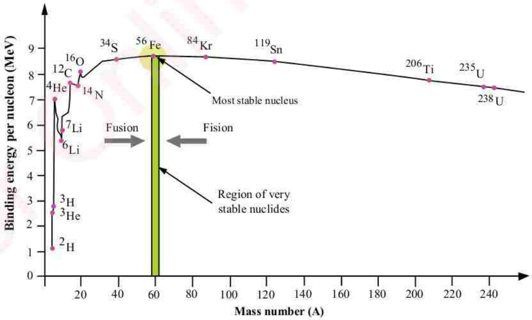


Figure 4.21 A variation of binding energy per nucleon with mass number

of the nucleus since the more nucleons the nucleus has, the greater the energy needed to separate them. Therefore, a more useful quantity for comparison is the binding energy per nucleon obtained by dividing binding energy (BE) by the nucleons number A. Graph of BE per nucleon against A is displayed in Figure 4.21. From this figure it is evident that atom with small and large nuclei are less bound than intermediate nuclei around ^{56}Fe. Relating this figure to nuclear binding energy will be useful during the discussion of nuclear stability in Section 4.4.3 and production of nuclear energy based on fusion and fission that will be presented in section 4.4.6.

4.4.3 Nuclear stability

From experimental observations some nuclides are stable in the sense that they remain unchanged indefinitely and some decay spontaneously to become other elements. The first category is called stable elements and the later are known as unstable or radioactive elements. There are two principal factors which can be used to establish the stability of a nucleus according to this definition.

The first is the binding energy per nucleon shown in Figure 4.21 and the second is the neutron to proton ratio depicted in Figure 4.22.

As described earlier in Figure 4.21, intermediate nuclei have the largest value of binding energy per nucleon around ^{56}Fe which has a value of $8.8\ \text{MeV / nucleon}$ as the most stable nucleus. Three nuclides $^{4}_{2}He$, $^{12}_{6}C$ and $^{16}_{8}O$ lie significantly above the main curve because of their high stability.

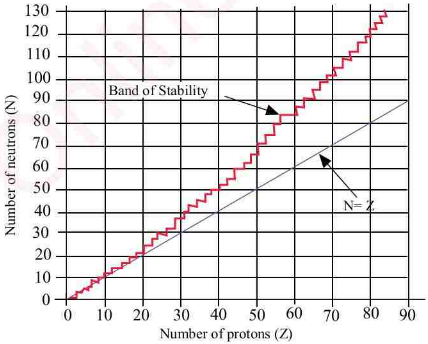


Figure 4.22 Variation of neutrons and protons numbers for known stable nuclei

From Figure 4.22, it can be seen that for the light nuclei the tendency is to have equal numbers of protons and neutrons up to about $A = 20$. Beyond this point the tendency is to have higher number of neutrons than protons because of increasing repulsive force among protons. This phenomenon explains the deviation of the stability curve from the straight line. For very heavy nuclei Z above 82, no number of neutrons can overcome the repulsive force from neutrons and therefore no stable nuclides exist with Z greater than 82.

If a nuclide has a combination of protons and neutrons which form an intersection outside the stability line, such nuclides is known to be unstable. There are three types of instabilities.

First the nuclide has the number of neutrons and protons that form an intersection above the stability line for low value of A. Such a nuclide tries to reach a stability line by emission of beta negative β^- and generally described by:

$$ {}_{Z}^{A}X \rightarrow {}_{Z+1}^{A}Y + {}_{-1}^{0}\beta + \bar{\upsilon} \qquad (4.39)$$

A specific example of this transformation is given by:

$$ {}_{6}^{14}C \rightarrow {}_{7}^{14}N + {}_{-1}^{0}\beta + \bar{\upsilon}$$

where $\bar{\upsilon}$ is called antineutrino with no charge and negligible rest mass.

Second, the nuclide has the number of neutrons and protons that form an intersection below the stability line for low value of A. Such a nuclide tries to reach a stability line by emission of beta positive β^+ and generally described by:

$$ {}_{Z}^{A}X_N \rightarrow {}_{Z-1}^{A}Y + {}_{+1}^{0}\beta + \upsilon \qquad (4.40)$$

A specific example of this transformation is given by:

$$ {}_{8}^{15}O \rightarrow {}_{7}^{15}N + {}_{+1}^{0}\beta + \upsilon$$

where υ is called neutrino with no charge and negligible rest mass. The transformation described by equations (4.39) and (4.40) are called *isobaric transformation;* they do not change the mass number.

Third, the nuclide has the number of neutrons and protons that form an intersection above the stability line for high value of A. Such a nuclide tries to reach a stability line by emission of an alpha particle α and generally described by:

$$ {}_{Z}^{A}X \rightarrow {}_{Z-2}^{A-4}Y + {}_{2}^{4}He \qquad (4.41)$$

A specific example of this transformation is given by:

$$ {}_{84}^{208}Po \rightarrow {}_{82}^{204}Pb + {}_{2}^{4}He$$

4.4.4 Radioactivity

Radioactivity is a random process by which a nucleus of an unstable atom loses energy by emitting nuclear radiations such as alpha particles, beta particles and gamma rays. Material consisting of atoms that spontaneously emits such radiations is said to be radioactive. Radioactivity as a random process at a level of single atom makes it difficult to predict when a specific atom will decay. However, when we have a collection of radioactive atoms, the decay rate is proportional to the number (N) of radioactive atoms present in a sample. Experimentally, the rate of decay is directly proportional to the number of radioactive nuclei present at that particular time and since radioactive decay decreases this number, the rate of decay can be given by:

$$-\frac{dN}{dt} = \lambda N \qquad (4.42)$$

where λ is called the *decay constant*. The minus sign indicates the number of radioactive atoms decreases with time. Re-arranging the terms of equation (4.42) we get:

$$\frac{dN}{dt} = -\lambda N \qquad (4.43)$$

Assuming the number of radioactive atoms at time $t = 0$ is N_o, then equation (4.43) can be integrated using these limits as follows:

$$\int_{N_o}^{N} \frac{dN}{N} = \int_0^t -\lambda\, dt$$

$$N = N_o e^{-\lambda t} \qquad (4.44)$$

where N is the amount of a radioactive material remaining after time t. If we define activity as a product of N and λ then, equation (4.44) can also be written as:

$$A = A_o e^{-\lambda t} \qquad (4.45)$$

where A is the activity of the material at time t and A_o is the initial activity. The SI unit of activity is *Becquerel (Bq)* or one disintegration per second. To describe radioactivity, scientists prefer to use half-life $t_{1/2}$ instead of the decay constant λ. As shown in Figure 4.23, *half-life* is defined as the time required for the radioactivity of a sample to reduce to half its initial value. The half-life $t_{1/2}$ is therefore a better term than the decay constant λ for differentiating among radioactive substances.

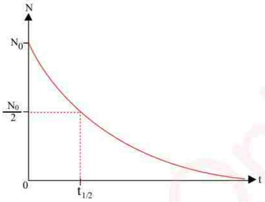


Figure 4.23 Radioactive decay curve

According to equation (4.44) and Figure 4.23:

$$\frac{N_o}{2} = N_o e^{-\lambda t_{1/2}} \qquad (4.46)$$

Simplifying equation (4.46) and rearranging terms gives:

$$t_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \qquad (4.47)$$

Example 4.14

A sample of radioactive material contains of 10^{18} atoms. The half – life of the material is 2000 days. Calculate:

(a) the fraction remaining after 5000 days; and

(b) the activity of the sample after 5000 days.

Solution

(a) The remaining fraction is;

$$\frac{N}{N_o} = e^{-\frac{\ln 2}{2000} \times 5000}$$

$$\frac{N}{N_o} = 0.18$$

(b) From part (a) above;

$N = 0.18N_o = 0.18\times10^{18}$

By definition $A = \lambda N$

$$A = \frac{\ln 2}{2000\times24\times3600}\times\left(0.18\times10^{18}\right)$$

$A = 722\,\text{MBq}$

Example 4.15

A sample of radioactive material has an activity of $9\times10^{12}\,Bq$. The material has a half-life of 80 seconds. How long will it take for the activity to fall to $2\times10^{12}\,Bq$?

Solution

Equations (4.45) and (4.47) give:

$$A = A_o e^{-\frac{\ln 2}{t_{1/2}}t}$$

$$2\times10^{12} = 9\times10^{12}\left(e^{-\frac{0.693}{80}t}\right)$$

$$\ln\left(\frac{9}{2}\right) = \frac{0.693t}{80}$$

$t = 174\text{s}$

Example 4.16

The half-life of strontium-90 is 28.8 years. Find:

(a) its decay constant in seconds;

(b) the initial activity of 4 g of strontium;

(c) how much activity will be remaining after 4 half-lives?

Solution

(a) $$\lambda = \frac{0.693}{t_{1/2}} = \frac{0.693}{28.8\times365\times24\times3600\text{s}}$$

$\lambda = 7.63\times10^{-10}\text{s}^{-1}$

(b) $A_o = \lambda N_o$

$$N_o = \frac{4.00\times10^{-3}\,\text{kg}}{90\left(\frac{\text{u}}{\text{atom}}\right)\times1.67\times10^{-27}\left(\frac{\text{kg}}{\text{u}}\right)}$$

$= 2.66\times10^{22}$ atoms

Therefore,

$A_o = 7.63\times10^{-10}\,\text{s}^{-1}\times2.66\times10^{22}$ atoms

$20.3\times10^{12} = 20.3$ TBq

(c) $A = A_o e^{-\lambda(4t_{1/2})}$

$$A = 20.3\,TBq\times e^{-7.63\times10^{-10}s^{-1}\times\left(4\times9.08\times10^{8}s\right)}$$

$A = 1.27$ TBq

4.4.5 Uses and hazards of radioisotopes

Uses

Nuclear radiations from radioisotopes are used in various fields including medicine, industry, research, and agriculture. Some of the applications are briefly explained below.

In medicine, radiation from radioisotopes is used for diagnosis or treatment of diseases. For example, radioactive iodine is used to treat cancer of the thyroid and Tc-99m is mostly used in medical imaging of different organs. In industry, radiation from radioisotopes is used in many areas such as smoke detectors, gauging thickness of paper, as tracers to locate leakage in pipes and assessment of road compactness during construction.

In research, radiation from radioisotopes is used to study properties of materials under different conditions. In agriculture it is used to study plant physiology and soil water content.

Hazards

Some of health effects induced by radiation in living organisms will be briefly discussed. When radiation from isotopes interacts with living cells, it can cause mutation. If the DNA in the nucleus of a cell is damaged by radiation, it may become cancerous. Cells die when they are exposed to a large amount of radiation, for example if the eye is exposed to radiation, the exposed person gets cataract. Exposure of the bone marrow to radiation lowers the immunity because the body fails to produce white blood cells in the bone marrow. When excessive exposure of radiation to an individual takes place a disease called leukemia or bone marrow cancer results.

4.4.6 Nuclear fusion and fission

It can be seen from the binding energy curve shown in Figure 4.21, that the binding energy per nucleon for light element is smaller; this means that fusing two elements to produce a heavier element would increase binding energy per nucleon. This binding energy appears as energy which could be utilized for generation of electricity. The process of producing energy by fusing two small nucleons is called *nuclear fusion*. This process is used to produce solar energy in the interior of the sun.

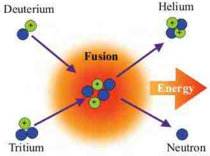


Figure 4.24 Fusion of tritium and deuterium

The fusion depicted in Figure 4.24 can be summarized by the equation;

$$ {}_{1}^{2}H + {}_{1}^{3}H \rightarrow {}_{2}^{4}He + {}_{0}^{1}n + Energy \quad (4.48)$$

It is interesting to note that in the nuclear reaction described by equation (4.48) the number of nucleons and charge are conserved. Using the concept of mass defect the energy released from fusion can be estimated as follows;

$$E = \begin{pmatrix} \left(M_{{}^2H} + M_{{}^3H}\right) - \\ \left(M_n + M_{He}\right) \end{pmatrix} \times 931.5 \frac{MeV}{u}$$

$$E = \begin{pmatrix} \left(2.014102\text{u} + 3.016049\text{u}\right) - \\ \left(1.008665\text{u} + 4.002603\text{u}\right) \end{pmatrix} \times 931.5 \frac{\text{MeV}}{\text{u}}$$

Therefore, energy released by fusing tritium and deuterium is $E = 17.59\ \text{MeV}$

Similarly, from Figure 4.21, the binding energy per nucleon for heavy elements is smaller than for medium elements. This means that, breaking a heavy element to produce two intermediate elements with increased overall binding energies also produces energy which could be utilized for generation of electricity. The process of producing energy by breaking heavy nuclei as illustrated in Figure 4.25 is called *nuclear fission*. It is interesting to observe that, the fission of U-235 described in Figure 4.25 produces different fragments including the ones described by the following equations;

$$\begin{aligned} &{}_{0}^{1}n+{}_{92}^{235}U \rightarrow {}_{56}^{141}Ba+{}_{36}^{92}Kr+3{}_{0}^{1}n+Energy \\ &{}_{0}^{1}n+{}_{92}^{235}U \rightarrow {}_{54}^{140}Xe+{}_{38}^{94}Sr+2{}_{0}^{1}n+Energy \quad (4.49) \\ &{}_{0}^{1}n+{}_{92}^{235}U \rightarrow {}_{50}^{132}Sn+{}_{42}^{101}Mo+3{}_{0}^{1}n+Energy \end{aligned}$$

As before the number of nucleons and charge in the nuclear reactions (fission) are conserved.

Again, using the concept of mass defect, the energy released from fission can be estimated as follows;

$$E=\begin{pmatrix}\left(M_{{}^{1}n}+M_{{}^{235}U}\right)- \\ \left(M_{{}^{141}Ba}+M_{{}^{92}Kr}+3M_{{}^{1}n}\right)\end{pmatrix}\times 931.5\frac{\text{MeV}}{\text{u}}$$

$$E=\begin{pmatrix}\left(1.008664u+235.043929u\right)- \\ \begin{pmatrix}140.914411u+91.926156u \\ +3.025992u\end{pmatrix}\end{pmatrix}$$

$$\times 931.5\frac{\text{MeV}}{\text{u}}$$

Therefore, energy released by fission of U-235 is $E = 173.29\ \text{MeV}$

This energy is in the form of kinetic energy of the fission products or fragments and the released neutrons. Neutrons released from fission are called high energy neutrons with typical energies of about 2 MeV. The high energy neutrons are not as effective as thermal neutrons in producing fission.

In view of fission efficiency of thermal neutrons, it is useful to estimate the energy of thermal neutrons. This is the kinetic energy a neutron would have at room temperature (300 K). Thus, thermal energy is of the order;

$$K_{Th}=\frac{3}{2}KT=\frac{3}{2}\times\left(8.617\times10^{-5}\,\text{eV/K}\right)\times\left(300\,\text{K}\right)$$

$$K_{Th}=0.039\,\text{eV}\approx 0.04\,\text{eV}$$

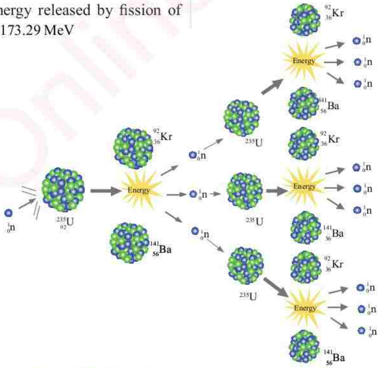


Figure 4.25 Nuclear fission and chain reaction of U-235 atoms

4.4.7 Operation of a nuclear reactor

Chain reaction

As already described in previous section, a slow neutron induces fission of $^{235}_{92}U$ thereby producing three neutrons and energy as an initial step towards the production of energy in the form of heat in a nuclear reactor. The three neutrons produced, often called first generation neutrons, which induce further fission reaction to produce second - generation neutrons as shown in Figure 4.25. The repetition of this process is called *chain reaction*.

Nuclear reactor

The first nuclear reactor based on the chain reaction described above was built by Enrico Fermi in 1942. A nuclear reactor is a power plant that produces electricity from the energy released by fission of radioactive materials such as U-235.

If one neutron produces about 200 MeV, a large number of neutrons produced in the chain reactions can produce substantial energy in a form of heat in the reactor core as shown in Figure 4.26. In order to increase the fissions efficiency the neutrons must be thermalized. This is achieved by collision of high energy neutrons with hydrogen component of water in the reactor core. The rate of production of neutrons in the reactor is controlled by insertion of the cadmium control rods (neutron absorbers). This allows the rate of production of neutrons in the chain reaction to match the rate required to produce the rated power of the plant by fission. The control rods are therefore used to ensure that the power plant produces heat from fission at desired rate. The water in the primary loop is pressurized to ensure that the temperature $\left(400^{o}C-500^{o}C\right)$ is sufficiently high to heat water and produce steam in the second loop at lower temperature $\left(100^{o}C\right)$. Like any other thermal power plants (fossil fuels) the steam generated at high pressure generates electricity by using a turbine connected to a generator. Once the steam passes the turbine it has to be cooled by water in the tertiary loop connected to the cooling tower.

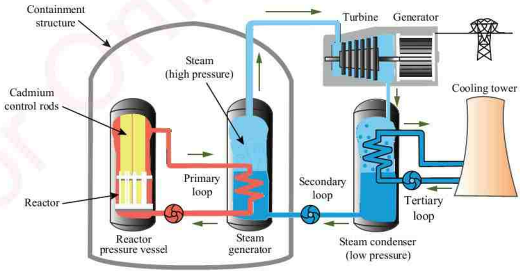


Figure 4.26 Nuclear reactor based on U-235 to produce electricity by fission

The cooling process condenses the steam to produce liquid water that can be re-used to produce more steam and the process continues. In the case of an accident radioactivity (fission fragments) released from the reactor core is confined by containment structure to minimize environmental pollution. It is important to observe that, to minimize environmental contamination, the water in the three loops circulates in isolation.

Essential requirements for nuclear power plant in Tanzania

As already described above, the core of the nuclear power plant acts like a furnace in which uranium $^{235}_{92}U$ acts like fuel. To produce energy by this furnace at a constant rate without exploding three important complications must be overcomed.

(i) Natural uranium found in Tanzania consists of 0.7%$\left(^{235}_{92}U\right)$and 99.3% $\left(^{238}_{92}U\right)$; the later does not produce energy in the core reactor. To increase the probability of neutrons produced to induce the next fission the percentage of $\left(^{238}_{92}U\right)$ must be reduced to 96%. The process used to do so is called natural uranium *enrichment*. Thus, nuclear fuel for power plant is a mixture of about 5%$\left(^{235}_{92}U\right)$and 95%$\left(^{238}_{92}\mathrm{U}\right)$.

(ii) The second requirement is to thermalize high energy fission neutrons using hydrogen in water in the first loop as a moderator.

(iii) The third requirement is to reduce the escape of the neutrons produced from the core reactor by increasing the mass of nuclear fuel used in a reactor. The minimum mass of uranium needed to reduce loss of neutrons to the level that can sustain chain reaction is called a *critical mass*.

It is useful to compare features of controlled fission in a power plant and un-controlled fission in an atomic bomb. As the word implies, un-controlled chain reaction proceed to an explosion while controlled fission produces energy at desired constant rate. To ensure that small amounts of weapon fuel is used to enable planes to deliver the bomb to the enemy, enrichment of uranium should be increased so that over 90% is $\left(^{235}_{92}U\right)$and 10% is$\left(^{238}_{92}U\right)$. This explains why nations that have signed the Non-Proliferation treaty are not allowed to possess high enriched uranium.

Example 4.17

(a) How much $^{2}_{1}H$ in kg/s is required to generate 100 MW by fusion according to equation (4.48)?

(b) Calculate the energy in joules, released from 50 kg of U-235; assume energy released by fission is 173.29 MeV

Solution

(a) According to equation (4.48) the energy produced by fusion is:

$$E = 17.59\,\text{MeV} \times 1.602 \times 10^{-13}\frac{\text{J}}{\text{MeV}}$$

$$= 2.8179 \times 10^{-12}\,\text{J},$$ then we have;

$$\text{Mass} = (2.014102\,\text{u}) \times (1.67 \times 10^{-27})\frac{\text{kg}}{\text{u}}$$

$$= 3.3636 \times 10^{-27}\,\text{kg} \qquad (i)$$

$$\text{Since power (P)} = 10^{8}\,\frac{\text{J}}{\text{s}} \qquad (ii)$$

It follows, $\frac{m}{s} = \frac{P}{E} \times \text{m}$

Applying cross multiplication in equations (i) and (ii) we get:

$$\frac{m}{s} = \frac{3.3636 \times 10^{-27}\, kg \times 10^{8}\, \frac{\text{J}}{\text{s}}}{2.8179 \times 10^{-12}\ \text{J}}$$

Therefore mass per second (m/s)

$$= 1.19 \times 10^{-7}\, \frac{\text{kg}}{\text{s}}$$

Energy production by fusion is very efficient because you need very mall amount of fuel to produce huge amount of energy.

(b) The energy in joules, released from 50 kg of U-235 is estimated as follows:

$$(235.043929\, u) \times \left(1.67 \times 10^{-27}\, \frac{kg}{u}\right) \equiv$$

$$173.29\ MeV \times \left(1.602 \times 10^{-13}\, \frac{J}{MeV}\right) \qquad (iii)$$

$$50\, kg \equiv x \qquad (iv)$$

Solving for x by cross multiplication of equations (iii) and (iv) we get;

$$x = \frac{50\, kg \times 2.7761 \times 10^{-11}\ \text{J}}{3.9252 \times 10^{-25}\ \text{kg}}$$

$$= 3.536 \times 10^{15}\ \text{J} = 3536\ \text{TJ}$$

Therefore, energy released from 50 kg of U-235 is 3536 TJ

Exercise 4.4

1. What are nuclides $^{7}_{3}X$ and $^{4}_{3}Y$ called? Which of the two is more likely to be unstable? Why?
2. (a) Calculate the mass defect when:
   (i) A neutron and a proton combine to form a deuterium nucleus.
   (ii) Two neutrons and two protons combine to form alpha particle.

   (b) Use the mass defect obtained in part (a) *i* and *ii* to calculate binding energy of deuterium and alpha particle in MeV.
3. Use mass energy equivalence ($E = mc^2$) to show that $1u = 931.5\, MeV$
4. Calculate the radioactivity in Bq of one gram of $^{226}_{88}Ra$ if its half-life is 1622 years.
5. Briefly explain how average binding energy or binding energy per nucleon and the neutron/proton ratios are used to:
   (a) Differentiate between nuclear fission and nuclear fusion.
   (b) Identify spontaneous emission of β^- from the nucleus.
   (c) Identify spontaneous emission of β^+ from the nucleus.
6. How much energy is needed to break the $^{12}_{6}C$ nucleus into three nuclei of $^{4}_{2}He$ or alpha particles? Take mass of carbon and alpha particles to be $12.000000u$ and $4.002603u$ respectively.
7. Estimate the energy Q that will be released in the following fusion reactions:
   (a) $^{1}_{1}H + ^{1}_{0}n \rightarrow ^{2}_{1}H + Q_1$
   (b) $^{2}_{1}H + ^{2}_{1}H \rightarrow ^{4}_{2}He + Q_2$
   (c) $^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{2}He + ^{1}_{0}n + Q_3$
8. $^{235}_{92}U$ undergoes a series of a combination of alpha and beta negative β^- decay to become a stable $^{238}_{92}Pb$. How many alpha and beta particles are emitted in the decay series?

9. The binding energy per nucleon of $^{235}_{92}U$ is about 7.5 MeV while the binding energy per nucleon for nuclei half the mass of $^{235}_{92}U$ is 8.5 MeV. Estimate using the value stated, the amount of energy released by breaking $^{235}_{92}U$ into two fragments.
10. Calculate the binding energy of $^{126}_{52}Te$ if the mass of this element is 125.903322 u.

Revision Exercise

1. Explain why classical physics fails to explain the following experimental observations of the photoelectric effect.
   (a) Kinetic energy of emitted electrons is independent of intensity of radiation.
   (b) Electron emission is spontaneous.
2. Explain how Einstein theory resolves the two failures of classical physics mentioned in question (1).
3. Why alkali metals e.g. K and Na, are most suited for photoelectric emission?
4. Explain, using Einstein theory, why photoelectric emission does not take place if the frequency of the incident photons is below a threshold frequency.
5. Why is the threshold frequency of sodium different from that of copper?
6. Radio Tanzania broadcasting station transmits at 103.7 MHz with power output of 200 kW. How many photons are emitted by the station per seconds?
7. If the wavelength of a photon that just separates a diatomic molecule is 300 nm, what is the binding energy of the molecule?
8. Light of wavelength 450 nm is incident on two photoelectric tubes. Suppose the threshold wavelength of the first tube is 600 nm and work function ϕ_1 is half the work function ϕ_2 of the second tube.
   (a) Find the stopping potential of each tube.
   (b) How do you explain the cause for the value of potential V_{o2} obtained for the second tube?
9. In a photoelectric experiment performed using calcium as an emitter, the following threshold frequencies f_o and stopping potentials V_o were recorded in Table 4.3. Use the information from the Table 4.3 to find the Planck's constant.

Table 4.3 Threshold frequencies and stopping potential

$f_o \times 10^{15} (Hz)$	$V_o (V)$
1.18	1.95
0.958	0.98
0.822	0.50
0.741	0.14

10. (a) The wavelength of light in a photoelectric experiment is increased from 300 nm to 303 nm. What is the corresponding change in the stopping potential?
    (b) Find the maximum kinetic energy of an electron emitted

from a surface with a threshold wavelength λ_c of 600 nm when light of wavelength 400 nm strikes the surface.

11. (a) What is the work function of a metal if the largest wavelength for photoelectron emission is 562 nm?
    (b) Suppose the metal is now illuminated with light of wavelength 250 nm. What will be the maximum speed of emitted electrons?
    (c) Suppose the intensity of light used in part (b) is $2\ \mathrm{W/m^2}$. What is the electron emission rate per unit area of the metal?
12. A monochromatic light of wavelength 450 nm falls on the metal surface of work function $\phi_o = 2.3$ eV. What is the value of:
    (a) the energy of the incident photons?
    (b) the maximum kinetic energy K_{max} of the electrons?
    (c) the threshold wavelength λ_o for the metal?
13. An electromagnetic radiation of wavelength 4×10^{-7} m is used to liberate electrons from a metal surface whose work function is 1.25 eV. Calculate the stopping potential for the metal.
14. A source of light is placed at a distance of 50 cm from an electron source. How much will the following quantities change if the distance is reduced to 20 cm? Use equations to explain whenever it is convenient.
    (a) Photoelectric current I
    (b) The threshold frequency f_o
    (c) Retarding potential V_o
    (d) The threshold wavelength λ_o
    (e) The maximum kinetic energy of the emitted photon K_{max}.
15. Explain, using an equation, why the increase of frequency of incident light on a metal has no effect on produced current from photoelectrons.
16. Explain using an equation why an increase of photon intensity produces an increase in current produced by photoelectrons.
17. Explain why, if the frequency and the light intensity are constant, the current decreases to zero as the retarding voltage is increased, when a certain stopping potential acquires a value V_o.
18. It is easier to remove an electron from sodium than from copper. Which of these has a higher value of threshold wavelength?
19. How would you convince your fellow student that wave-particle duality exists for all particles?
20. Why is the wave nature of particles not observed in normal life? Explain using a familiar particle travelling at normal speed.
21. At what speed would a bullet of mass $m_b = 10\,g$ travel for its wavelength to be comparable to that of an electron of mass m_e travelling at a speed 10^6 m/s? What conclusion can you make about a bullet?

22. (a) Using $E = \frac{3}{2}KT$, estimate the de Broglie wavelength of neutrons at room temperature 27 °C, where K is the Boltzmann's constant and T is the temperature.
    (b) Estimate the de Broglie wavelength of fission neutrons with kinetic energy of 8 MeV.
    (c) Explain why fast neutrons from fission (nuclear reactor) of energy of about 8 MeV must be thermalized (kinetic energy reduced to $E = \frac{3}{2}KT$) before they are used in neutron diffraction experiments.
23. Show that de Broglie hypothesis of matter wave is in agreement with Bohr's theory.
24. A proton is accelerated from rest through a potential difference of 1 kV. What is the de Broglie wavelength?
25. How did Rutherford interpret the observation that very few alpha-particles were deflected, by as large angle as 180^0, by the gold foil?
26. How did Rutherford interpret the observation that most of the alpha particles passed through the gold foil?
27. What is the difference between excitation energy and ionization energy? When are they the same?
28. What are the main differences between the Rutherford's model and the Bohr's model of the atom?
29. Explain two features that make the Bohr's model of the atom superior to the model proposed by Rutherford. ***Hints***: orbits and stability.
30. Explain using a sketch diagram how de Broglie theory explains the Bohr's postulate of existence of stable orbits.
31. Explain how Rutherford's experiment on scattering of alpha particles with gold foil can be used to estimate the nuclear radius of the gold atom.
32. (a) What is the accelerating potential that will enable an electron in singly ionized helium to get into the second excited state?
    (b) Use Bohr's theory to estimate the second ionization of the singly ionized helium.
33. What is the highest state that hydrogen atoms at ground state can reach when they are bombarded by electrons accelerated by a potential difference of 12.6 volts?
34. In a transition to a state of excitation energy 10.19 eV a hydrogen atom emits a 489 nm photon. Determine the binding energy of the initial state.
35. Laser light results from transition from long-lived metastable stages. Why is it more monochromatic than ordinary light?
36. (a) Explain the production of laser light.
    (b) What is meant by population inversion?
    (c) Mention properties of laser light.
    (d) Explain the applications of laser light in different areas.

37. A laser emits light with a frequency of 4.69×10^{14} Hz.
    (a) What is the energy of one photon of the radiation from this laser?
    (b) If the laser emits a pulse of energy containing 5.0×10^{17} photons of this radiation, what is the total energy to the pulse?
    (c) If the laser emits 1.3×10^{-2} J of energy during a pulse, how many photons are emitted during the pulse?
38. How many protons and neutrons are in element ${}^{232}_{92}Y$?
39. (a) If the radius of a nucleus is $R = R_o A^{\frac{1}{3}}$ and its mass is A, show that nuclear density is independent of A.
    (b) Use $R = R_o A^{\frac{1}{3}}$ to estimate the mass number of a stable nucleus whose radius is $\frac{1}{3}$ that of ${}^{189}O_s$.
40. The density of copper is about $10^5\ \text{kg/m}^3$ but the density of its nucleus is about $10^{17}\ \text{kg/m}^3$. Why are the atomic and nuclear densities of the same element so different?
41. Use the stability line to explain why ${}^{12}_{7}N$ is radioactive. What decay process of ${}^{12}_{7}N$ is more likely in its attempt to become a stable element? What stable element will be formed after decay?
42. Someone claimed that hydrogen ${}^{1}_{1}H$ has no nuclear binding energy. Is this claim true or false? Discuss.
43. What is the difference between the elements that decay by positron and electron emission? Give an example of each.
44. (a) Given mass of the proton, $M_p = 1.00728u$ and mass of an electron, $M_e = 0.00055u$, what would be the mass of the hydrogen atom?
    (b) Calculate the binding energy per nucleon for ${}^{4}_{2}He$ atom given that;

    $$M_n = 1.00867u,\ M_p = 1.00728u$$
    $$M_e = 0.00055u,\ M_{He} = 4.00260u$$

    (c) The binding energy per nucleon for ${}^{12}_{6}C$ is 7.68 MeV and that for ${}^{13}_{6}C$ is 7.47 MeV. Calculate the neutron separation energy of ${}^{13}_{6}C$, i.e. energy required to remove a neutron from ${}^{13}_{6}C$.
45. What does it mean when one says that the half-life of a radioactive sample is 6 hours?
46. (a) There are 10^8 radioactive atoms in a sample. If its half-life is 40 s, how many radioactive atoms will be left after 20 s?
    (b) The half-life of radium is 1590 years. How long will it take in years to reduce 1g of radium to 1 centigram?
47. In an experiment to determine the half-life of a radioactive material a student measured the radioactivity of a sample after every of 5minutes interval and obtained the data depicted in Table 4.4.

Table 4.4 Time (minutes) and activity (MBq)

Time (minutes)	Activity (MBq)
0	19.2
5	7.13
10	2.65
15	0.99
20	0.37

(a) Use the data to determine the half-life of the radioactive sample.

(b) How much of radioactive material will be present in the sample after 120 minutes from the time the student started doing the experiment?

48. In measurement of radioactive decay by Geiger-Muller counter the following count rates were obtained at the times shown in Table 4.5. Use a semi-log plot to determine the half-life of the radionuclide.

Table 4.5 Time (hours) and count rate (cpm)

Time (hours)	Count rate (cpm)
0	4032
1	3075
2	2341
3	1780
4	1375
5	802
6	619
7	475
8	371
9	287

49. Why is high temperature required to induce nuclear fusion in collection of hydrogen atoms?

50. Assuming that the sun radiates energy like a blackbody, estimate the rate at which it is losing its mass per second.

    Given that the radius of the sun is 696392 km, surface temperature of the sun is 5890 K and Stefan's constant $\sigma = 5.6704x10^{-8}\,\mathrm{W/m^2K^4}$.

51. A nuclear reactor is a device in which controlled chain reaction takes place to produce heat for electricity generation. The essential parts of a nuclear reactor are

    (i) Nuclear fuel

    (ii) Moderator

    (iii) Control rods

    (iv) Coolant

    (v) Protective shield.

    Explain the function of each part.

52. How is controlled nuclear chain reaction achieved in a nuclear power plant based on the nuclear fission of $^{235}_{92}U$?

53. (a) Determine the value of y in the following nuclear fission reaction of $^{235}_{92}U$ atom using slow neutrons:

    $$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{143}_{56}Ba + ^{90}_{36}Kr + y^{1}_{0}n + Energy$$

    (b) Explain why for chain reaction to take place y must be greater than 2?

54. Why do we need a cooling loop of water in a fission based on nuclear reactor?

55. In fission, $^{235}_{92}U$ releases 173 MeV. What is the fission rate required to produce power of 1MW ?

Chapter Five

Environmental Physics

Introduction

Environmental processes in our everyday life require an understanding of the interaction between organisms and their environment. The interrelationships between the atmosphere, biosphere, hydrosphere and lithosphere are included in the environmental processes. These processes embody very important concepts such as solar radiation, heat transfer, and the flow of water in soil which are useful in agriculture. Moreover, it is important to understand the conversion of solar, wind, geothermal, and ocean wave energy into electrical energy. In this chapter, you will learn about the agricultural physics, energy from the environment, earthquakes, and environmental pollution.

5.1 Agricultural physics

Agricultural physics deals with agricultural materials and physical processes aimed at improving the quantity and quality of agricultural products, and taking into account the role of environmental factors. Applying the principles of physics in agriculture such as heat transfer and flow rate of water in the soil enables informed choice of plants to be grown. In this section description is based on the influence of radiation budget in the environment, soil properties and elements of weather on plant growth.

5.1.1 Solar radiation

Solar radiation refers to electromagnetic radiation emitted by the sun. The sun emits radiation approximately as a blackbody (energy emitted per unit area $= \sigma T^4$) at a temperature of about 6000 K. Most of the emitted radiation has wavelengths between 0.2 μm to 2.0 μm, with the maximum around 0.5 μm and the visible radiation has wavelengths between 0.4 μm and 0.7 μm Figure 5.1.

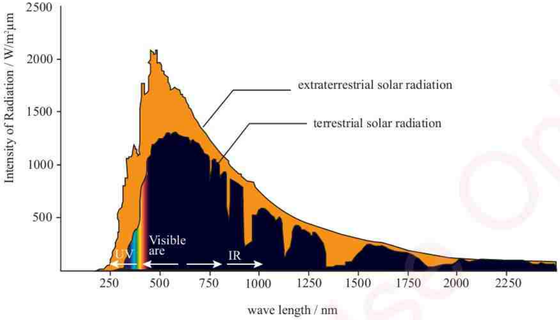


Figure 5.1 The terrestrial (at sea level) and extraterrestrial (above the earth's atmosphere) solar spectrum

At the top of the atmosphere direct solar beam can be considered to be formed of parallel rays. The irradiance of a surface normal to the beam *(extraterrestrial radiation)* at the top of the atmosphere is about 1380 Wm^{-2}. The direct solar beam is attenuated as it passes through the atmosphere due to absorption and scattering Figure 5.2, so that the irradiance at the earth's surface directly overhead in cloud-free conditions is reduced to about 1000 Wm^{-2}.Some of the radiation scattered from the direct beam by molecules and aerosols reaches the earth's surface as diffuse solar radiation.

It should be noted that the earth's surface and the atmosphere emit electromagnetic energy referred to as terrestrial radiation by virtue of their temperature. The Earth's surface emits like a blackbody at a temperature of about 290 K. From Wien's displacement law $\left(\lambda_{\max} \propto \frac{1}{T_s}\right)$ the wavelength range of terrestrial radiation will be much larger than that of solar radiation. Most of the energy is in the wavelength range 3 µm to 30 µm, with a maximum around 10 µm. Terrestrial radiation is often referred to as long-wave radiation, while extraterrestrial is referred as shortwave. For most surfaces the long-wave radiation emitted per unit area is given by $\varepsilon_s \sigma T_s^4$, where ε_s is an effective emissivity of the surface.

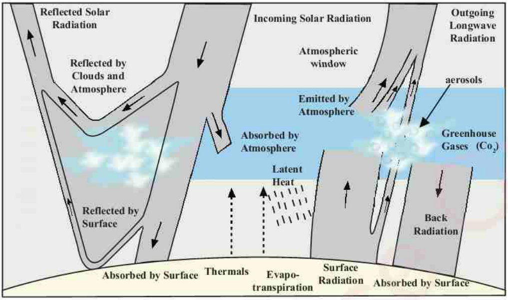


Figure 5.2 Earths energy budget

From Kirchhoff's law, the surface absorbs the same amount of radiation as it emits. The emissivity of the atmosphere varies strongly with wavelength, so the atmosphere does not behave like an ideal blackbody. For wavelengths between 8 µm and 12 µm the atmosphere is almost transparent so the emissivity is very small. This wavelength range is known as *atmospheric radiation window*. For wavelengths between 3 µm and 8 µm and for wavelengths greater than 12 µm the atmosphere is nearly opaque, i.e. the emissivity is close to one. The emission and absorption characteristics of the atmosphere varies with the concentrations of green-house gases such as water vapour, methane and carbon dioxide. Variations in water vapour are responsible for the short term variability in the radiative characteristics of the atmosphere. Liquid water has an emissivity close to one, so that thick clouds act like blackbody radiators. The greenhouse gases absorb long wave radiations emitted by the earth's surface then re-emit long wave radiation back to the earth surface.

For the earth's temperature to remain constant the incoming and outgoing energy should be balanced Figure 5.2 and therefore the net solar irradiance (S_n) of the surface can be represented as:

$$S_n = S\downarrow - S\uparrow \qquad (5.1)$$

where $S\downarrow$ is the total incoming short-wave radiation and $S\uparrow$ is the total outgoing short-wave radiation. Since the earth surface does not emit short-wave radiation $S\uparrow$ is entirely associated with reflection of some of the down-welling radiation. It follows from equation 5.1 that:

$$S_n = (1-\alpha)\times S\downarrow = (1-\alpha)\times(S_b + S_d) \qquad (5.2)$$

where $\alpha = \dfrac{S\uparrow}{S\downarrow}$ is the surface albedo, S_d is the direct-beam solar irradiance and S_d is the diffuse solar irradiance whose magnitude depends on atmospheric aerosol and clouds.

The outgoing long wave radiation ($L\uparrow$) from the earth's surface is

$$L\uparrow = \varepsilon_s \sigma T_s^4 + \left(1-\varepsilon_s\right)L\downarrow \qquad (5.3)$$

The first term on the right-hand side of equation (5.3) is the radiation emitted by the earth's surface and the second term is the reflected down-welling long wave radiation incident on the surface. The net long wave irradiance (L_n) of the surface is

$$L_n = L\downarrow - L\uparrow = \varepsilon_s \times \left(L\downarrow - \sigma T_s^4\right) \qquad (5.4)$$

In the field the net long wave irradiance will be measured directly and the earth's surface temperature will also be estimated. With these measurements equation (5.4) can be used to estimate the downward long wave irradiance ($L\downarrow$) as:

$$L\downarrow = \sigma T_s^4 + \frac{L_n}{\varepsilon_s} \qquad (5.5)$$

The net surface irradiance R_n is

$$R_n = S_n + L_n \qquad (5.6)$$

The net solar radiation is very important for growth and development of plants, specifically the visible portion for green plants to transform light energy into chemical energy by a process of photosynthesis. The infrared radiation is responsible for the heating effect which result into the temperature of the environment, which is one of the primary factors affecting the plant growth.

5.1.2 Influence of aerial environment on plant growth

Aerial environment refers to the diverse atmospheric conditions and processes in the atmosphere in terms of vapour pressure, humidity, temperature and wind.

(a) Vapour pressure and humidity

Air contains water vapour which can be either saturated or partially saturated. Saturation conditions exist when the air contains the maximum possible level of water vapour for existing temperature conditions. Generally, the air is partially saturated when the density of water vapour in the air is less than the maximum possible level. Vapour pressure (e), which is the partial pressure exerted by the water vapour, is more commonly referred to equillibrium pressure of vapour and is related to vapour density and temperature (T) by the ideal gas law. The saturated vapour pressure is always measured at the ambient air temperature (e_a) and at the dew point temperature (e_d) and the two are related to relative humidity (RH) as:

$$RH = \frac{e_d}{e_a} \times 100\,\% \qquad (5.7)$$

The amount of water moisture in the air (humidity) is measured in terms of the saturated vapour pressures at ambient and dew temperatures. It is also important to note that as plants transpire, the humidity around them saturates the leaves with water vapour. When relative humidity levels are too high, a plant cannot make water evaporate or draw nutrients from the soil. When this occurs for a prolonged period, a chance for plant diseases increases.

(b) Air temperature

Air temperature influences the rate of physical and chemical processes that determine plant's growth and development. These processes include photosynthesis, breaking of seeds, seed germination, translocation, protein synthesis, respiration and transpiration. Each plant has its own optimum temperature range for growth and development. Some plants grow better in cooler temperatures while others prefer warmer temperatures. The growth rate of most plants increases as the temperature increases provided that the optimum temperature for enzymes activity is not exceeded. High air temperature generally increases the loss of moisture from the soil and from plants. Ambient air temperature is the most common temperature reported, and it is measured using a standard thermometer. Wet bulb and dew point temperatures characterize the moisture properties of the air.

(c) Wind

Wind plays an important role as a forced vapour removal mechanism. Immediate vapour removal around a plant creates a large vapour pressure gradient and this enhances evaporation from surface and evapotranspiration from plant leaves. This can lead to water deficit to plants and hence affect plant growth. On the other hand if wind is too low, moisture accumulates and vapour pressure increases. Such condition is favourable to plant growth but it can also create conducive condition for development of plant diseases. Wind is responsible for transport pollutant soil dust particles which can eventually settle on plant leaves. Dust cover can lead to reduction in light uptake by plants and so reduction in photosynthesis, hence reduction in dry matter production. Strong wind can also damage plants by uprooting or breaking them.

5.1.3 Soil

Soil is the loose surface material consisting of mineral, water, gases, organic matter and micro-organisms. Soil provides structural support to plants, and is also a source of mineral nutrients and water. Physical and chemical properties of soil vary greatly, depending on their ages and conditions (climate, vegetation, parent materials and topography) in which it was formed. Processes such as weathering, leaching, and microbial activities also influence soil properties.

(a) Soil components

Soil is made up of solid, liquid and gaseous phases. The solid phase has two main components which are mineral (inorganic) and organic soil.

(i) The mineral or inorganic component consists of particles of varying sizes, shapes, and chemical compositions, mostly a product of weathering of the parent rock that include quartz, feldspar, magnetite, garnet, silicates and other secondary minerals. It is the largest component of soil of up to about 50% by volume. The particles have densities ranging between 2000 kg/m^3 – 2800 kg/m^3.

(ii) The organic component includes plant and animal residues in different stages of decomposition as well as active living organisms. It makes up to about 5 % of the total soil volume, with particles density of 1200 kg/m^3 to 1500 kg/m^3. This component has large surface area and high capacity to hold water as well as other essential elements important for plant growth. Micro-organisms are found in very large numbers as part of the organic component of soil but make as low as about 1% of the soil volume. A variety of chemical and physical factors are affected by microbes. These include, but not limited to texture, temperature, pH, oxygen, cation exchange capacity and redox reactions. The soil environment affect directly the types of microbes as well as the rate of processes they perform. For example, microbial activities increase with temperature, which in turn affects the rate of decomposition.

(iii) The liquid phase consists of a dilute aqueous solutions of inorganic and organic compounds which fill up part or all of the open spaces between the solid particles. It is the second basic component accounting for about 2% to 50% of soil volume. The chemical composition and the freedom with which it moves varies depending on the other soil components and soil texture.

(iv) The gaseous phase is composed of soil air which occupies the part of the pore space between the soil particles that is not filled with water. Since air dissolves in water, it coexists with water making approximately 2% to 5% of soil volume. Oxygen in soil is essential for roots and CO_2 and N_2 are also needed to support plant growth and plant functions through processes such as nitrogen– fixation, photosynthesis and the carbon cycle.

(b) Soil properties

The composition and proportions of components of soil discussed in the previous sub section greatly influences soil physical properties such as structure, texture, soil temperature and heat, soil water and soil aeration. Soil structure is the arrangement and organization of soil particles and the way individual soil particles bind together to form aggregates. Aggregation creates pore space which is instrumental for flow and retention of water and gases. It has been noted that effects of soil structure on plant growth includes, water and nutrients uptake by plants, microbial activities and root growth.

The chemical properties of soils that are important to plant growth are:

(i) Nutrient availability and cation exchange capacity, which collectively affect the soil's inherent fertility and its ability to hold nutrients.

(ii) The chemical characteristics of the soil solution, which affect pH and salinity.

(c) Soil structure

Soil structure can be classified in terms of the shape that the soil takes based on its physical and chemical properties. Each individual unit of undisturbed soil structure is called a *ped.* Various types of peds are described as follows:

Granular: These are usually less than 0.5 cm in diameter and are commonly found in surface horizons where roots have been growing.

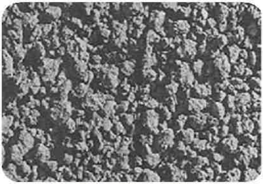


Figure 5.3 Granular shaped soil

Blocky: Irregular blocks that are usually about 1.5cm - 5.0 cm in diameter.

Figure 5.4 Blocky shaped soil

Prismatic: Vertical columns of soil that might be some few centimeter long. Usually found in lower horizons.

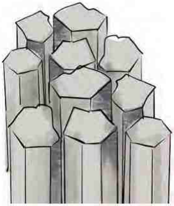


Figure 5.5 Prismatic shaped soil

Columnar: Vertical columns of soil that have a salt "cap" at the top. Found in soils of arid climates.

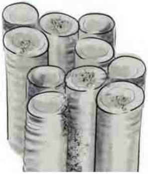


Figure 5.6 Columnar shaped soil

Platy: Thin, flat plates of soil that lie horizontally. Usually found in compacted soil.

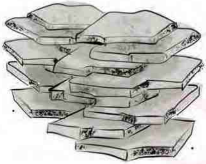


Figure 5.7 Plate shaped soil

Single Grained: Soil is broken into individual particles that do not stick together. Commonly found in sandy soils.

Figure 5.8 Sandy soil

Massive: Soil has no visible structure, it is hard to break apart and appears in very large clods

Figure 5.9 Massive blocks of soil

(d) Soil texture

Soil texture is the relative proportions of sand, silt and clay particles in a mass of soil. Many of the important soil properties are directly influenced by soil texture, however the most important effect on plant growth is water holding capacity and consequentially nutrient supply and uptake. The proportions are determined using the textural triangle method based on the United State Department of Agriculture (USDA) system of particle size as shown in Table 5.1 and Figure 5.10.

The soil textural triangle is a tool that can be used to visualize and understand the meaning of the soil texture names. It is a diagram in which 12 soil textures are classified based on the percentage of sand, silt and clay in each soil sample. The system defines soil particle as sand if the size is 0.1 mm to 2 mm, silt when all particles are within size range of 0.002 mm to 0.05 mm and clay for all particles with size less than 0.002 mm.

Table 5.1 *USDA soil texture classes*

Common name of soil (General texture)	Sand (Percentage weight)	Textual class
Sandy soils (coarse texture)	70-86	Loamy sand
Loamy soils (moderately coarse texture)	50-70	Sandy loam
Loamy soils (medium texture)	23-52	Loam
	20-50	Silty loam

To determine soil texture using the textural triangle, the following steps are used

(a) Find the percent of clay, silt and sand in your sample using sieve method.

(b) Find and locate the percentage of sand along the base of the triangle and draw a line going up towards the left.

(c) Find and locate the percentage of clay along the left side of the triangle and draw a line towards the right until it meets the previous line for sand. The point of intersection is the sample texture.

(d) Find and locate the percentage of silt along the right side of the triangle and draw a line down until it meets the point of intersection between clay and sand.

(e) Determine the soil sample texture by the area of the triangle in which the intersection point is located.

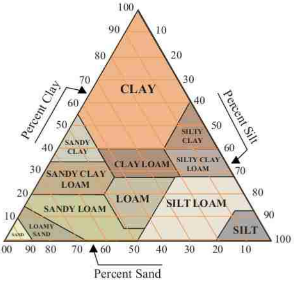


Figure 5.10 Triangle diagram of the basic soil textural classes.

Example 5.1

Suppose a particular soil sample is comprised of 30 % clay, 60 % silt and 10% sand. Determine the textural class of the soil.

Solution

To determine the textural class of the soil sample shown in Figure 5.11, use the following steps:

(a) We are provided with soil sample comprised of 30 % clay, 60 % silt and 10% sand.

(b) We locate the 10 % sand along the base of the triangle in Figure 5.11 and then draw a line (blue) along the 10 % line for sand until it meets the 30 % clay .

(c) We locate the 60 % silt on the right of the triangle and draw a line (green) downward to the left until it meets the line for 10 % sand.

(d) The point of intersection between red, blue and green lines falls under the silty clay loam textural class (enclosed by the purple lines) and this is the textural class of the soil provided sample.

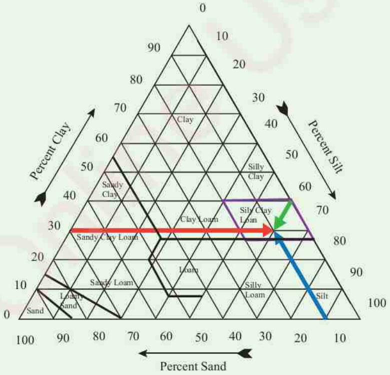


Figure 5.11 Determining the textural class of the soil sample with 30 % clay, 60 % silt and 10%

Activity 5.1

Identify the texture of a soil sample using the triangle diagram of the textural classes

(e) Water flow in soil

There are two important forces which influence water flow in soil, namely, gravitational and surface tension forces. These forces can be expressed in terms of the soil-water potential, Ψ, which is defined as *the amount of work per unit quantity of pure water that must be done by external forces to transfer reversibly and isothermally an infinitesimal amount of water from the standard state to the soil at the point under consideration*. The volume of water that can flow through a unit cross-sectional area of the soil, Q_W can be determined by the Darcy's Law, written as:

$$Q_W = -\kappa \frac{d\Psi}{dx} \qquad (5.8)$$

The negative sign indicates that water flows from high to low potential. The constant κ is the effective permeability of the soil also known as the hydraulic conductivity; it is a measure of the soil's ability to permit water to flow through its pores or voids. The effective permeability κ depends on the size and distribution of pores. The top soil is often moist due to the balance between gravity and suction which ensures that there is some moisture to be returned to the atmosphere through evaporation.

Example 5.2

A sample of loam sand soil is tested in a laboratory to measure its hydraulic conductivity as shown in Figure 5.12. The column has an inside diameter of 18 cm and the length between manometers is $\Delta s = 28\,\text{cm}$. With a steady flow of 1.9 cm³/min the head difference between the manometers $\Delta h = 14\,cm$. Calculate the hydraulic conductivity.

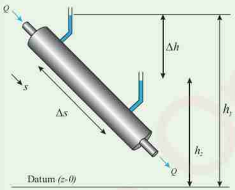


Figure 5.12 Steady flow through a sand sample

Solution

Using the Darcy's law for one-dimensional flow,

$$Q_W = -\kappa \frac{d\Psi}{dx} = -\kappa \frac{dh}{ds} A$$

Where, A is the cross sectional area of the manometers, Q_W is the discharge rate and κ the hydraulic conductivity in the s direction. Then

$$\kappa = -Q_W \frac{1}{A}\frac{ds}{dh}$$

$$= \left(-1.9\,\text{cm}^3\,\text{min}^{-1}\right) \times \left(\frac{1}{\pi(9\,\text{cm})^2}\right)\left(\frac{28\,\text{cm}}{-14\,\text{cm}}\right)$$

$$\kappa = 0.015\,cm/\text{min} = 2.5 \times 10^{-4}\ \text{cm/s}$$

The sign of $\frac{ds}{dh}$ is negative because as s increases, h decreases.

Activity 5.2

Compare the flow rate of water through different types of soil.

(f) Heat transfer in soils

Soil temperature is important in determining the rates and directions of soil physical and chemical processes as well as mass and energy exchange with the atmosphere. Temperature also influences biological processes in the soil such as seed germination, root growth and microbial activity. All mechanisms of heat transfer, i.e. conduction, convection and radiation may take place, however the primary one for soil is conduction. An important parameter for measuring the ability of the soil to transfer heat by conduction is the specific heat capacity defined as the amount of heat required to raise the temperature of one kilogram of a substance by one degree. The heat capacity C of soil is given as:

$$C = c_{soil} \times mass \qquad (5.9)$$

where c_{soil} is the specific heat capacity of the soil

$$c_{soil} = x_s c_s + x_w c_w + x_a c_a \qquad (5.10)$$

where x_s , x_w and x_a are the soil volume fractions of the solid material, water and air with their corresponding specific heat capacities c_s ,c_w , and c_a, respectively. In one dimension heat conduction K is described by a diffusion equation:

$$\frac{K}{C}\left(\frac{d^2T}{d^2z}\right) = \frac{dT}{dt} \qquad (5.11)$$

where $\frac{dT}{dt}$ is the rate of change of temperature with time and $\frac{d^2T}{d^2z}$ is the spacial change of temperature gradient and K is thermal conductivity. The one dimensional soil heat flux flow per unit area per unit time (G) for a homogeneous medium is given by the Fourier's law of heat conduction as:

$$G = -K\frac{dT}{dz} \qquad (5.12)$$

Heat flow in soil facilitates temperature distribution in the soil and has direct and indirect effect of plant growth and development as well as the biological, chemical and physical processes in the soil. The importance of equations (5.11) and (5.12) will be illustrated using worked examples.

Example 5.3

If the temperature difference across a 43 cm thick dry sand soil is 23 °C, calculate the thermal flux and total heat transfer in three hours under steady state conditions. Assume thermal conductivity of dry sand soil as $0.27\ \mathrm{W/m^oC}$.

Solution

The heat flux across a sand soil column is expressed as

$$G = -K\frac{d\left(T_1 - T_o\right)}{dz} =$$

$$\frac{\left(0.27\,\mathrm{W/m^oC}\right)\times\left(23^oC\right)}{\left(0.43\,\mathrm{m}\right)} = 14.4\,\mathrm{Wm^{-2}}$$

The total heat transfer $= G \times t$

$$= \left(14.4\,\mathrm{Js^{-1}m^{-2}}\right)\times\left(3\times3600\mathrm{s}\right)$$

$$= 1.6\times10^5\ \mathrm{Jm^{-2}}$$

Example 5.4

You are provided with a soil column containing 53 cm of dry sand over 29 cm of dry loam soil. The ends of dry sand and dry loam soils are attached to constant temperature baths with the top

Physics Form VI X.indd 225 10/11/2019 17:24

maintained at 37°C and the bottom at 14°C. Given that thermal conductivity of dry sand soil is $0.27\ \text{W/m}^{o}C$ and that of dry loam soil is $0.15\ \text{W/m}^{o}\text{C}$, calculate:

(a) The steady state heat flux through the two layers

(b) The temperature at the sand-loam interface.

Solution

(a) Let K_S, K_L and K_{eq} be the thermal conductivity for sand soil, loam soil and equivalent thermal conductivity of the dry sand-loam soil. Also the thickness of sand, loam and the total thickness of the dry sand-loam system are donated by z_s, z_l and z_T respectively.

From equation (5.12)

$$G = -K\frac{dT}{dz} \Rightarrow \frac{dz}{K} = \frac{-dT}{G}$$

At the sand - loam interface $\frac{dT}{G}$ is the same and constant and hence the value of K_{eq} due to total thickness z_T is given as

$$\frac{z_T}{K_{eq}} = \frac{z_s}{K_s} + \frac{z_l}{K_l} \Rightarrow K_{eq} = \frac{z_T}{\left(\frac{z_s}{K_s}\right) + \left(\frac{z_l}{K_l}\right)}$$

substituting the values of z_s, z_l, K_s and K_l then,

$$= \frac{0.82\text{ m}}{\left(1.96\,m^2\,{}^{\circ}\text{CW}^{-1} + 1.93\text{m}^2\,{}^{\circ}\text{CW}^{-1}\right)}$$

$$K_{eq} = 0.21\text{W/m}\,{}^{\circ}\text{C}$$

The heat flux across the dry sand-loam soil column is given as

$$G = -\frac{K_{eq}dT}{z_T}$$

$$= -\frac{0.21\text{W}/m\,{}^{\circ}\text{C}}{\left(0.53\text{m} + 0.29\text{m}\right)}\left(37 - 14\right){}^{\circ}\text{C}$$

$$= -\frac{0.21\text{W/m}^{o}C \times 23^{o}\text{C}}{0.82\text{ m}}$$

$$\therefore G = -5.89\text{ Wm}^{-2}$$

(b) The temperature T_i across the sand-loam interface is given by

$$G = -\frac{K_S}{z_s}\left(37^{\circ}C - T_i\right) \Rightarrow T_i = 37^{\circ}C + \frac{Gz_s}{K_s}$$

$$T_i = 37^{o}C + \frac{\left(-5.89\text{ Wm}^{-2}\right)\left(0.53\text{ m}\right)}{0.27\ \text{W/m}^{o}C}$$

$$T_i = 37^{o}C - 11.56^{o}C$$

$$T_i = 25.44^{o}C$$

5.1.4 Techniques for improving plant environment

Techniques used for improving plant environment are mulching, shading and sheltering.

(a) Mulching

Mulching is a mechanism of modifying the properties of a soil by placing any material on soil surface such as sawdust, manure, straw, leaves, crop residue, gravels, paper, and plastic sheets. This technique improves plant environment in several ways.

(i) Reducing evaporation and soil erosion

Paper or plastic mulches, particularly the light colored ones, are effective in reducing soil evaporation while vegetative mulch must have sufficient thickness to be effective in reducing evaporation and soil erosion from rain Figure 5.13.

Figure 5.13 Plastic mulching

(ii) ***Controlling weeds***

Weeds need light to grow, applying mulch such as plastic mulches blocks access to sunlight and consequently degrades weed growth.

(iii) ***Regulating and moderating soil temperature***

Soil thermal profile is a function of the coefficient of thermal conductivity and heat capacity of the soil. A mulch of dry crop residue as shown in Figure 5.14 provides lower coefficient of thermal conductivity and consequently reduces temperature variations in the soil.

Figure 5.14 Dry crop mulching

(iv) Other benefits to plant growth include increasing population of micro-organism, improving soil structure, providing organic matter and reduce soil borne diseases.

(b) Shading

Shading is a process of protecting plants from excessive solar radiation as shown in Figure 5.15. This process helps to prevent excessive loss of water in plants through transpiration. Shading also changes other environmental conditions such as temperature, humidity and carbon dioxide concentrations.

Figure 5.15 Plant shading

(c) Sheltering

Sheltering is a mechanism of protecting plants from the adverse effect of strong winds by reducing the wind speed. Strong winds can physically damage plants, causing soil erosion and consequently hinder the overall plant growth. One method of reducing the wind speed is the use of windbreaks. Windbreak is the form of a shelter belt usually made up of one or more rows of trees or shrubs planted in such a manner to provide shelter to plants from the wind as shown in Figure 5.16. Through their ability to reduce speed, wind breaks increase soil temperature and humidity as well as reduce soil erosion.

Figure 5.16 Wind break

Activity 5.3

Design a small garden using techniques for improving plant environment.

Exercise 5.1

1. Why would life on earth be impossible if there was no greenhouse effect.
2. What is meant by the term *earth's energy balance?*
3. Discuss the importance of solar radiation on plant's growth and development.
4. Describe the four major components of soil.
5. Describe the importance of shading, mulching and sheltering in agriculture.
6. A soil column has a 2 mm thick crust at the top and a total length of 20 cm, thus the remaining length of column is 19.8cm. Water is kept ponded on the surface to a depth of 1 cm and steady-state flow of water takes place through the column; the column is open to the atmosphere at the bottom. The saturated K of the crust is $0.001\,\text{cm/hr}$ and the hydraulic conductivity of the underlying soil is $5\,\text{cm/hr}$.
   (a) What is the effective hydraulic conductivity of the column? What is the flux density?
   (b) What is the pressure potential on weight basis at the interface between the crust and the underlying soil layer?
7. What is meant by soil permeability and soil porosity?
8. A soil sample 42 cm thick has a temperature difference of 13°C across the column. Calculate:
   (a) the thermal flux; and
   (b) the total heat transfer in one hour under steady state condition.
   (Use: Thermal conductivity of soil $= 1.6\,\text{W/mK}$.)
9. A farmer experiences surface soil temperature variation during daily activities. Explain the factors which influence the surface soil temperature.
10. Discuss the greenhouse effect and its importance in daily life experience in plant growth.

5.2 Energy from the environment

The sources of energy in our environment can be classified as renewable and non renewable sources. The later is not replenished in nature, therefore the focus has been on renewable energy.
The renewable energy resources available in our environment include solar radiation, wind, tides, waves, bioenergy and geothermal energy. These energy resources

are renewable because they are replenished by nature within a reasonable short time.

5.2.1 Solar energy

Most of the energy used today originates from the sun. Plants require energy from the sun for growth, water power depends on the hydrological cycle that is facilitated by solar energy. Wind energy requires air pressure differences that are provided by unequal heating of the land and water bodies by the sun. Even fossil fuels (non renewable) such as coal, oil and natural gas have their origin in the heat provided by the sun. Probably only nuclear and geothermal energy are not directly connected to the sun. People use solar energy to produce electricity as well as thermal energy (drying crops, salt extraction, solar cooker).

(a) Solar photovoltaics

Solar radiation energy can be converted to electricity by photoelectric effects, usually using semiconductors in a device called a solar cell, Figure 5.17 (a). The current used materials for photovoltaics include monocrystalline silicon, polycrystalline silicon, amorphous silicon, cadmium telluride, and copper indium gallium selenide/sulfide. A solar cell is, in principle a p-n junction operated in a reverse bias. A p-n junction is created by bridging a p-doped semiconductor and n-doped semiconductor. The holes from the p-type side will diffuse to n-type side while the electrons will diffuse to the p-type side, creating, a junction with a fixed negative charge in the p side and fixed positive charge in the n side as illustrated in Figure 3.9 on chapter three.

The fixed charges create an electric field which makes it easy for current to flow in one direction but hard to flow in the opposite direction.

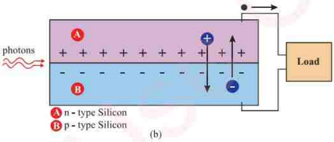


Figure 5.17 (a) Solar cell and (b) photovoltaic cell

When photons from the sun fall on the junction Figure 5.17(b) electrons are exited and if the photon energy is equal or greater than the semiconductor band gap energy E_g, electrons can jump from the valence band to the conduction band leaving behind holes in the valence band as shown in Figure 5.18.

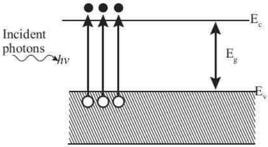


Figure 5.18 Photon induced electron-hole generation in a semiconductor.

The inbuilt electric field across the p-n junction separates electrons and holes that have been created by the absorption of sun light. When the electrons and holes are separated electric power can be extracted from the circuit. Only light with energy greater than the bandgap (conduction valence bands separation) of a semiconductor can create (through excitation) free electrons or holes that can be collected into a useful electricity as shown in Figure 5.19.

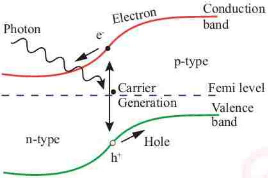


Figure 5.19 Electron-hole pair separation by the inbuilt electric field

The current I generated by the solar cell is a combination of the photo generated current I_L and the *dark* forward bias diode current I_o is given by:

$$I = I_o\left[e^{\frac{qV}{kT}} - 1\right] - I_L \qquad (5.13)$$

where, V is the cell voltage, T is the cell temperature, k is Boltzmann's constant and q is charge. The I-V characteristics of solar cell can be obtained experimentally by varying the load R in Figure 5.20 and recording the corresponding current and voltage.

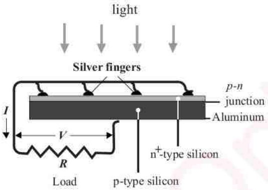


Figure 5.20 Solar cell circuit

A typical silicon solar cell generates an open circuit voltage, V_{oc}, ranging from about 0.5 to 0.7 V. The current generated depends on the size of the surface area of the cell. To have a usable electricity (voltage and current), solar cells are connected in series and parallel configurations into modules and arrays in solar photovoltaic systems. *(a group of solar cells are called modules and a group of modules are called arrays)* for example a module designed to charge a 12 volt battery, 36 cells are connected in series to get an open circuit voltage between 18 to 21 V required for the given task (note that a solar cell is not operated at the open circuit level). If higher currents are needed, solar cells or modules can be connected in parallel.

Example 5.5

A solar cell of surface area 100 cm^2 is illuminated by a monochromatic light of a wavelength of 760 nm and power density of 1000Wm^{-2}. The open circuit voltage is 0.67 V when the cell temperature is 300 K. Calculate the dark cell current I_o, assuming a 100 % quantum efficiency.

Solution

The current generated by the solar cell is given by:

$$I = I_o\left[e^{\frac{qV}{kT}} - 1\right] - I_L$$

Open circuit voltage is defined when the current $I = 0$; the above therefore equation becomes

$$0 = I_o\left[e^{\frac{qV_{oc}}{kT}} - 1\right] - I_L \Rightarrow e^{\frac{qV_{oc}}{kT}} = \frac{I_L}{I_o} + 1$$

For 100 % quantum efficiency, $\frac{I_L}{I_o} >> 1$ and hence $e^{\frac{qV_{oc}}{kT}} = \frac{I_L}{I_o}$ or

$$V_{oc} = \frac{kT}{q}\ln\left(\frac{I_L}{I_o}\right)$$ evaluating for $\frac{I_L}{I_o}$

$$\frac{I_L}{I_o} = e^{\frac{qV}{kT}} \Rightarrow \frac{I_L}{I_o}$$

$$= e^{\left(\frac{1.6\times10^{-19}\ C\times0.67\ V}{1.38\times10^{-23}\ JK^{-1}\times300\ K}\right)} = 1.75\times10^{11}$$

Power P intercepted by the cell is

$$P = 1000\ Wm^{-2}\times100\times10^{-4}\ m^2 = 10\ W$$

The frequency f of the monochromatic light is; $f = \frac{c}{\lambda}$, where c is the speed of light in free space.

$$f = \frac{3\times10^8\ ms^{-1}}{760\times10^{-9}\ m} = 3.95\times10^{14}\ Hz$$

The photon flux is

$\phi = \frac{P}{hf}$, where h is plank's constant.

$$= \frac{10\ Js^{-1}}{6.62\times10^{-34}\ Js\times3.95\times10^{14}\ s^{-1}}$$

$$\phi = 3.82\times10^{19}\ \text{photons}\ s^{-1}m^{-2}$$

With 100 % quantum efficiency, each photon generates 1 free electron, hence:

$$I_L = q\phi$$
$$= 1.6\times10^{19}\ C\times3.82\times10^{19}\ \text{photons}\ s^{-1}m^{-2}$$
$$= 6.11\ A$$

From $\frac{I_L}{I_o} = 1.75\times10^{11}$

$$\Rightarrow I_o = \frac{I_L}{1.75\times10^{11}}$$

$$I_o = \frac{6.11\ A}{1.75\times10^{11}} = 3.49\times10^{-11}\ A$$

The dark current is 3.49×10^{-11} A

(b) Standard illumination conditions

Power output and the efficiency of a solar cell are evaluated under the standard conditions of solar irradiance of 1000 W/m², at air mass (AM) 1.5 spectrum and ambient temperature of 25 °C. When the Sun is at the zenith under reasonable cloudless atmospheric conditions, the absorption due to the atmosphere is defined as air mass. In most cases, the zenith angle θ of the Sun is not zero as shown in Figure 5.21. The American Society for Testing and Materials (ASTM) chose as the standard condition when the absorption is 1.5 times the normal air mass that is 1.5 (AM).

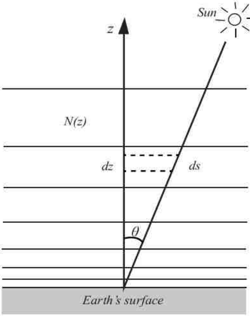


Figure 5.21 The sun at an angle θ from the zenith axis Z

It is noted from Figure 5. 21 that vertical elemental distance dz is given by

$$dz = ds\cos\theta \qquad (5.14)$$

where the curvature of the atmosphere has been ignored. The AM can be defined in terms of the vertical elemental distance dz, actual distance travelled by the sunlight ds and the zenith angle as:

$$AM = \frac{ds}{dz} = \frac{1}{\cos\theta} \qquad (5.15)$$

The standard zenith angle is therefore defined as:

$$\theta = \cos^{-1}\left(\frac{1}{AM}\right) = \cos^{-1}\left(\frac{1}{1.5}\right) = 48.19^{\circ} \qquad (5.16)$$

(c) Fill Factor of a solar cell

Fill factor is a measure of the quality of the solar cell and is calculated by taking the ratio of the maximum power the cell can generate to the product of open circuit voltage and short circuit current. The open-circuit voltage V_{OC} of a solar cell is the voltage provided by a solar cell under standard illumination conditions when the cell is not connected to any load the corresponding current being zero. The short-circuit current I_{SC} is the current of a solar cell when the load applied has zero resistance and the corresponding voltage is zero. For any finite load R, the voltage V and the current I will be smaller than the open circuit voltage and the short circuit current, respectively. The maximum power output of the solar cell can be determined from the condition as:

$$dP = d(IV) = IdV + VdI = 0 \qquad (5.17)$$

From Figure 5.22, the maximum power point is denoted by point (V_{mp}, I_{mp}). The fill factor of a solar cell is then defined as;

$$FF = \frac{Area\ A}{Area\ B} = \frac{P_{max}}{I_{sc}V_{oc}} = \frac{I_{mp}V_{mp}}{I_{sc}V_{oc}} \qquad (5.18)$$

where A represents area of the I-V curve when the circuit is closed and B the area of the I-V curve when the circuit is both open and closed

The higher the fill factor the higher the quality of the solar cell and vice versa.

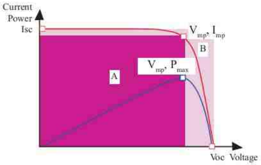


Figure 5.22 I-V and P-V curves for a typical solar cell

(d) Efficiency of solar cell

The *efficiency* of a solar cell is defined as the ratio of the output electric power to the input solar radiation power under standard illumination conditions at the maximum power point. The efficiency η of a solar cell is given by the ratio between the maximum power delivered by the cell, P_{max}, to the power P_r from the solar radiation intercepted by the cell at standard illumination conditions and is given by:

$$\eta = \frac{P_{max}}{P_r} \qquad (5.19)$$

Solar cell efficiency is usually limited by a number of factors including light absorption, charge separation, charge transport, cell temperature and internal resistance of the cell.

5.2.2 Wind energy

Wind is atmospheric air in motion caused by uneven absorption of solar radiation by the earth's surface. As a consequence of variation of heat capacity, earth's surface heats up and cools down unevenly, creating atmospheric pressure zones making air masses flow from high to low pressure zones. A good example of this phenomenon is sea and land breezes that, air flows from sea to land during the day and from land to sea during the night respectively. To harness energy from the wind we need a device to convert the kinetic energy of the wind into usable form of energy such as electrical or mechanical energy by a turbine. Wind turbines, convert the kinetic energy of the wind into rotational kinetic energy that can be useful for a variety of purposes. The main uses of a wind turbine is generation of electricity and for pumping water from boreholes. Other uses include grinding and engine for sailing boats.

(a) Wind turbines

Usually the rotational mechanical energy in the turbine is converted by a generator into electrical energy. These turbines are classified as either small scale which produce less than 300 kW while large turbine produce greater than 300 kW. A typical wind turbine for generation of electricity is as shown in Figure 5.23.

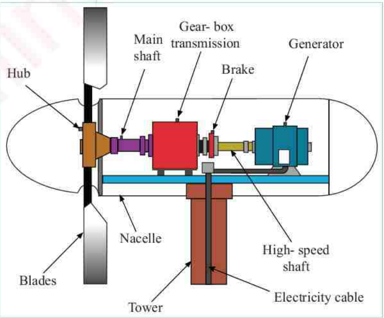


Figure 5.23 Components of a wind turbine for generation of electricity

The generator is connected to the turbine shaft through a gearbox transmission which turns it at a different speed usually higher than the main turbine shaft. The blades operate on either of the two principles. The first principle is based on drag where the wind pushes the blades sideways. Force act in the direction of wind. The second is the principle of lift off where the blade is designed as an aerofoil so that when air flows past the blade, the wind speed on the lower surface of the blade is lower than that of the upper surface. Force is perpendicular to direction of wind. It is clear that the air pressure at the upper surface of the aerofoil is lower than that at its lower surface, hence the pressure difference is created between the surfaces creating a lift up which in turn, turns the turbine shown in Figure 5.24.

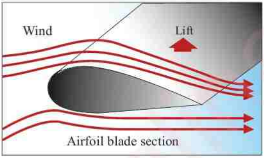


Figure 5.24 An aerofoil creating a lift

(b) Extractable power from a wind turbine

The power P in wind contained in a cylindrical column (that is intercepted by the horizontal blades) of free unobstructed wind moving horizontally at a constant speed v is equivalent to the rate of change of its kinetic energy E given as:

$$P = \frac{dE}{dt} = \frac{d}{dt}\left(\frac{1}{2}mv^2\right)$$

$$= \frac{1}{2}\left(2mv\frac{dv}{dt} + v^2\frac{dm}{dt}\right) \qquad (5.20)$$

For a constant wind speed, $\frac{dv}{dt} = 0$, and hence the power P becomes:

$$P = \frac{1}{2}\left(v^2\frac{dm}{dt}\right) \qquad (5.21)$$

If the cross-sectional area of the air column is A and its density is ρ then the mass flow rate $\frac{dm}{dt}$ is given as:

$$\frac{dm}{dt} = \rho A v \qquad (5.22)$$

The power in equation (5.21) becomes:

$$P = \frac{1}{2}\rho A v^3 \qquad (5.23)$$

Thus, wind power in an open air stream is proportional to the third power of wind speed. This means that when the wind speed is doubled the wind power increases eight times. Wind turbines for grid electricity therefore need to be efficient at a greater wind speed. Equation (5.23) assumes 100% energy conversion efficiency, however real turbines cannot convert kinetic energy of wind by more than 59.3 % into the mechanical energy by turning a rotor. This conversion efficiency limit is known as the Betz Limit or Betz' Law. The theoretical maximum power efficiency of any design of wind turbine is expressed in terms of *power coefficient* and is defined as $C_{p\max} = 0.59$.

The practical C_p value is dependent on the turbine type and is a function of wind speed. Hence, the power coefficient needs to be factored in equation (5.23) and the extractable power (P_{ext}) from the wind is given by:

$$P_{ext} = \frac{1}{2}\rho A v^3 C_p \qquad (5.24)$$

From equation 5.24 it is clear that power can be extracted from wind depending on environmental factors ρ and v as well as turbine design factors A and C_p. Since wind speed is influenced by terrain and height, these factors must be taken into consideration when installing a wind turbine for power generation.

(c) Wind turbine tip speed ratio

The Tip Speed Ratio (TSR) is an extremely important factor in wind turbine design. TSR refers to the ratio of the the speed of the tips of the wind turbine blades (rotor) to the wind speed.

$$\text{TSR} = \frac{\text{Tip speed of blade}}{\text{Wind speed}} = \frac{\omega r}{v}$$

If the wind turbine rotor spins too slowly, most of the wind will pass through the gap between the blades, providing no power to the blades. But if the rotor spins too fast, the blades will act like a solid wall to the wind, consequently no rotational motion. To get the maximum amount of power from the wind, the turbines must be designed with optimal tip speed ratios.

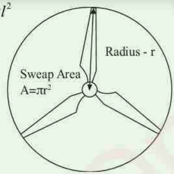
Example 5.6

Calculate power extractable by a wind turbine of blades' length 52 m if the wind is blowing at a speed of 12 m/s. Assume the density of air and power coefficient to be 1.23 kg/m³ and 0.4, respectively.

Solution

The area swept by blades of the turbine is calculated assuming blades sweeps circular area whose radius is equal to the blade length.

$$A = \pi r^2 = \pi l^2$$



$$P_{ext} = \frac{1}{2}\rho A v^3 C_p$$

$$= \frac{1}{2}\left(1.23\text{kgm}^{-3}\right)(\pi)(52\text{ m})^2(12\text{ m/s})^3(0.4)$$

$$P_{ext} = 3.61\text{MW}$$

Example 5.7

The rated output power for a turbine model at 15 m/s is 0.13 MW. If the rotor rotates at a constant frequency of 0.198 Hz, calculate:

(a) the tip speed to wind speed ratio given that the diameter of the rotor is 90 cm; and

(b) the power conversion efficiency (Density of air 1.225 kgm^{-3})

Solution

(a) The linear velocity of the tip:

$$v_t = \omega r = 2\pi f \frac{d}{2}$$

$$= 2\pi \times 0.198\text{ s}^{-1} \times \left(\frac{90\text{ cm}}{2}\right) = 0.56\text{ ms}^{-1}$$

The tip speed to wind speed ratio

$$TSR = \frac{v_t}{v_w} = \frac{0.56\text{m/s}}{15\text{m/s}} = 3.7\times10^{-2} = 0.037$$

(b) Wind power at 15 m/s

$$P = \frac{1}{2}\rho A v_w^3$$

$$=\frac{1}{2}\times\left(1.225\,\text{kg m}^{-3}\right)\times\pi\times\left(0.45\,\text{m}\right)^{2}\times\left(15\,\text{ms}^{-1}\right)^{3}$$

$$P=1.32\text{ kW}$$

The power conversion coefficient

$$C_{p}=\frac{1.32\text{ kW}}{0.13\times10^{3}\text{kW}}\times100\%=1.01\%$$

5.2.3 Geothermal energy

Geothermal energy is the energy obtained by tapping heat in the earth's interior. Part of this heat originates from the earth's interior due to the friction and gravitational pull which exist by virtue of Earth's composition, formation and the other part from decay of radioactive elements. Most of the earth's heat is generated through radioactivity processes deep into the core of the earth. The heat generated is radiated outward and if the temperature of the rocks reaches about 700°C – 1300° C, *magma* is formed.

When the magma heats nearby rocks and underground acquifers, hot water can be released through geysers, hot springs and steam vents. This heat can be extracted by drilling and injecting water to create steam. Hot water or steam generated in this process can be used for heating or directed as steam to a turbine to generate electricity. Most of the geothermal energy remains in the earth's mantle, slowly moving outwards collecting into high heat pockets. People use geothermal energy as a source of heat and electricity production.

(a) Extracting geothermal energy for heating and cooling

Geothermal energy can be used directly as a source of heat from low-temperature geothermal heat pockets (of about 150°C); these are usually found just a few metres below the ground. Geothermal heat pumps (GHPs) can be used to extract the earth's heat, through a pipe in a continuous loop that circulates water between the underground and above ground. Water or other liquid is pumped through the pipe during the cold season, it absorbs underground geothermal energy and carries it above the ground for use.

(b) Extracting geothermal energy for generation of electricity

Three types of power plants are used to generate electricity from geothermal energy, namely: Dry steam, flash steam, and binary cycle plants. Dry steam plants take steam out of fractures in the ground as shown in Figure 5.25 and use it directly to drive a turbine that spins a generator. Flash steam plants take hot water, usually at temperatures over 200 °C, out of the ground, and allow it to rise to the surface then separates the steam phase in steam/water separators and then runs it through a turbine. In binary plants, the hot water flows through heat exchangers, boiling an organic fluid that has a lower boiling temperature than water. The steam generated spins the turbine to generate electricity. It is expensive to build a geothermal power station but operating costs are low, resulting in low energy costs for suitable sites.

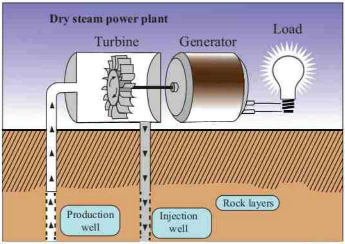


Figure 5.25 Dry steam geothermal plant

5.2.4 Tidal energy

Tidal energy is a form of hydropower that converts the energy carried by tides into electricity. Tidal power is predictable and depends on the gravitational pull of ocean water on earth by the moon and the sun. Creation of low and high tides depends on season and particular location on the earth. The basic principle beyond tidal power is to trap the tide behind a barrier during high tide and let the water out through a turbine at low tide.

If the tidal height is h and the entrapment area is A, the mass of water trapped is given as ρAh and since the centre of gravity is $\frac{h}{2}$ above the low tide level, then the maximum stored energy per tide is given by $(\rho Ah)(g)\left(\frac{h}{2}\right)=\frac{1}{2}\rho Agh^2$

If the tidal period is T, then the available power is:

$$\langle P \rangle = \frac{\rho Ah^2 g}{2T} \qquad (5.25)$$

The principle described above uses the potential energy difference between the height of the high and low tides by using a Tidal Barrage System as shown in Figure 5.26.

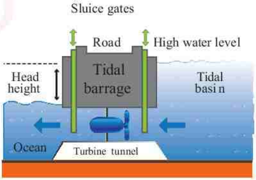


Figure 5.26 The tidal barrage system

Tidal Stream Systems is another method of extracting energy from tides; this method uses the kinetic energy of the moving water to rotate a turbine. Tidal stream systems can be placed where there are strong and fast currents. Good locations can be found at the entrances to bays and rivers and between islands or other land masses where the currents

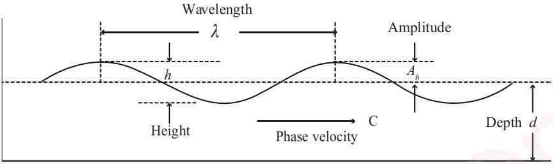


Figure 5.27 Water wave

become concentrated. Generating electricity from the tides in oceans is not widely used but it is an emerging marine renewable energy industry.

5.2.5 Ocean wave energy

When the wind blows over the surface of offshore water, such as oceans and seas, waves are created by the progressive transfer of energy from the wind. Wave energy is a stored concentrated wind energy. Water waves can be regarded as back and forth exchange between kinetic energy and potential energy of water. Figure 5.27 shows a sinusoidal surface wave which is symmetrical about the equilibrium position of the water level.

The total energy E in an ocean wave is the sum of the potential and kinetic energy; it can be expressed as:

$$E = \frac{\rho g A_b^2}{2} \text{ or } E = \frac{\rho g h^2}{8} \qquad (5.26)$$

where ρ is the water density, g is gravitational accelation, A_b is the wave amplitude, and h is the wave height.

Note: *Wave height h is twice the amplitude A_b.*

The average energy flux or power of a wave over the wave period T is given by multiplying energy E with its propagation speed $v_g = \frac{\lambda}{2T}$ and is given as:

$$< P > = \frac{\rho g A_b^2}{2}\left(v_g\right) = \frac{1}{2}\rho g A_b^2 \frac{\lambda}{2T} \qquad (5.27)$$

The dispersion relationship between period T and wavelength λ can be calculated using relation below:

$$\omega = \sqrt{gk} \Rightarrow \frac{2\pi}{T} = \sqrt{g\frac{2\pi}{\lambda}} \text{ or}$$

$$\lambda = \frac{gT^2}{2\pi} \qquad (5.28)$$

Then, substituting equation (5.28) into equation (5.27) gives:

$$< P > = \frac{1}{8\pi}\rho g^2 A_b^2 T = \frac{1}{32\pi}\rho g^2 h^2 T \qquad (5.29)$$

From equation (5.29) it is clear that, the higher the amplitude (wave height) of the wave, the more powerful the wave is. Wave power is determined by wave height, period (which is a function of wave speed), acceleration due to gravity and water density. Technologies for extracting wave energy have been designed and installed in the near shore (offshore) locations. There are four basic methods for extracting wave energy: Utilizing point absorbers, attenuators, overtopping devices, and terminators (oscillating water column) as described below.

(a) Terminators (oscillating water column)

A terminator device shown in Figure 5.28 is set perpendicular to the direction of the wave travel (indicated by the long horizontal arrow) to capture the power of the wave. In this device water enters into a column through a subsurface opening, trapping air above. The wave action moves the captured water column up and down as indicated by the vertical line with arrows on both ends. This in turn forces the trapped air through an opening connected to a turbine to generate power.

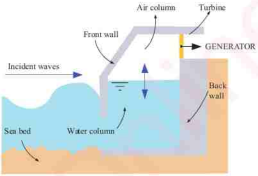


Figure 5.28 Oscillating water column that forces air through a generator

(b) Attenuators

Attenuators, also known as linear absorbers, are long horizontal semi-submerged floating structures oriented parallel to the direction of the waves as shown in Figure 5.29. A wave attenuator is made of a series of cylindrical sections linked by flexible hinges. The differing heights of waves along the length of the device causes different segments to rotate or yaw relative to each other. The wave-induced motion is used to pressurize a hydraulic piston which turns a hydraulic turbine generator to produce electricity.

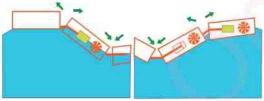


Figure 5.29 Attenuator in left side as the wave moves up and the right side as the wave moves down

(c) Point absorber

A point absorber is a device such as a floating buoy inside a fixed cylinder that moves relative to each other due to wave action. These devices convert the up and down pitching motion of the waves into oscillatory movement as indicated by arrows in Figure 5.30. The motion of the shaft drives an electricity generator.

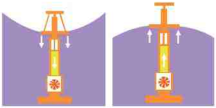


Figure 5.30 Point absorber device showing movement of bouy by arrows.

(d) Overtopping devices

Overtopping devices, also known as spill-over devices, are either fixed or floating structures with reservoirs that are filled with water by incoming waves as shown in Figure 5.31, causing a slight build-up of water as found in a dam. The potential energy of the trapped water in the reservoir is extracted by allowing the water to flow back to the sea through a turbine generator to produce electricity. The water is then released, and gravity causes it to flow back into the ocean. The energy of the falling water is used to turn hydro turbines to generate power.

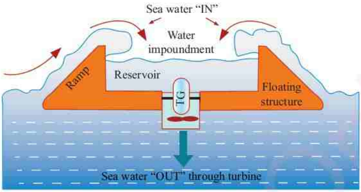


Figure 5.31 Overtopping device showing water into a reservoir by arrows

Special built floating platforms can also generated electricity by funnelling waves through internal turbines and then back into the sea as indicated by the arrow pointing vertically down in Figure 5.31.

Example 5.8

The city of Dar es Salaam sea shore is known to experience tidal ranges that can reach 4 m. Suppose a tidal barrage of 100 km² is constructed, how much total potential tidal energy can be reserved by the barrage in one year by using *bidirectional* turbines? Take the density of sea water and acceleration due to gravity to be 1023 kg/m³ and 9.8 m/s² respectively.

Solution

Lunar tidal period is 12.42 hours and number of tidal cycles per year is;

$$n_{cyc} = \frac{(24 \times 365.25)\text{ hours}}{12.42 \text{ hours}} = 705.80$$

Number of times the tide to drives the turbines becomes:

$$n = 2n_{cyc} = 2 \times 705.80 = 1411.6$$

Total tidal energy per year is therefore:

$$= \frac{nA\rho gh^2}{2}$$

By substituting the values

$$E = 2.83 \times 10^{15}\left(\frac{\text{kgm}}{\text{s}^2}\right)(\text{m})$$

$$= 2.83 \times 10^{15}\text{ Nm}$$

$$= 2.83 \times 10^{6} J \text{ (since } 1\text{ Nm} = 1\text{ J)}$$

$$\therefore E = 7.86 \times 10^{5}\text{ MWh}$$

Example 5.9

Calculate the wave power of a wave in a sea shore which move a height of 3 m. Take density of sea water 1023 kgm⁻³, and acceleration due to gravity 9.8 ms⁻².

Solution

The wave energy is given by,

$$E = \frac{\rho gh^2}{8}$$

$$E = \frac{1023\text{ kgm}^{-3} \times 9.8\text{ ms}^{-2} \times (3\text{ m})^2}{8}$$

$$E = 11278.6 \text{ J}$$

$$E = 11.28\text{ kJ (approx)}$$

Exercise 5.2

1. Briefly explain how wind power is generated.
2. Show that the total power generated by wind is given by $P = \frac{A}{2}\rho v^3$ where, A is area swept by blade through which air of density ρ passes and v is velocity of wind.
3. (a) What is meant by wind power? How is wind power captured by wind turbines?
   (b) Give at least three characteristics of wind power.
4. (a) What is wind turbine and how does it work?
   (b) What is meant by wind energy converter?
5. A wind turbine works in opposite to that of a fan. Justify your answer.
6. How does the horizontal wind turbine differ from vertical wind turbine.
7. What is meant by the following terms as regards to photovoltaics?
   (a) Open circuit voltage.
   (b) Short circuit current.
   (c) Maximum power point.
   (d) Fill factor.
   (e) Conversion efficiency.
8. Describe how a solar cell works.
9. The dark current density for a silicon solar cell at 35 °C is $1.6 \times 10^{-8}\ \text{Am}^{-2}$, voltage at maximum power is 0.549 V and the short circuit current density is $150\ Am^{-2}$. Calculate:
   (a) open circuit voltage;
   (b) current density at maximum power;
   (c) maximum power;
   (d) maximum efficiency; and
   (e) the cell area required for an output of 45 W when exposed to solar radiation of power density 950 Wm^{-2}.
10. Describe the design of three types of geothermal power plant.
11. Discuss any three human and environmental benefits of using wind energy.
12. A wind turbine made of blades of radius r, is driven by a wind of speed v. If ρ is the density of air, derive an expression for the maximum power, P, which can be produced by the turbine, in terms of ρ, r and v.
13. A water wave has the energy of 10000 J. Calculate the height of the wave.
14. Explain the meaning of the terms wave energy, tidal energy, and barrage.
15. Explain any three factors that hinder development of wave power in Tanzania.
16. Write down two sources of geothermal energy and tell what determines its conduction to the earth's surface.

5.3 Earthquakes

An earthquake is a sudden shaking or movement of the earth's surface ground, caused by the rupture or slippage of a fault within the earth's crust. An earthquake is one of the worst natural disasters that frequently happens in earthquake prone zones.

5.3.1 Elastic rebound theory

The elastic rebound theory describes the process of successive build-up of shear stress in a fault of a rock, release of strain energy from the rock in form of a stress wave followed by springing back of the rock. The theory is summarized as follows:

(a) Initially rocks A and B are separated by a fault (shown in Figure 5.32(a) as a dashed line). The two rocks try to move relative to each other because of the actions of opposing forces (indicated by arrows). However, this movement is prevented from happening by frictional force between the rocks thereby forming a *locked fault*. Prevention of the movement is indicated by the straight contours.

(b) As rock A and B on opposite side of a fault continue to be subjected to shear stress, they are strained, and therefore build up elastic energy which progressively deforms the rocks bend as indicated by the contours in Figure 5.32(b).

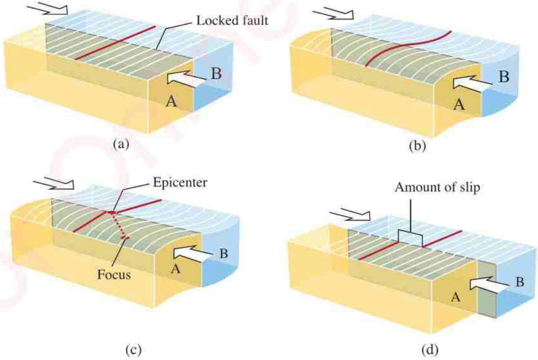


Figure 5.32 An illustration of elastic rebound theory

(c) When the shear stress exceeds the frictional force along the fault, rocks A and B suddenly slip and thereby suddenly release the stored energy as seismic waves (earthquake). The point where slipping of the rocks started is called focus (hypocentre). The point above the focus is called the epicentre as shown in Figure 5.32 (c).

(d) As soon as the stress is released, the rocks A and B spring back (rebound) to their original un-deformed shape as indicated by displaced straight contours in Figure 5.32(d). It is noted that the elastic rebound theory put forward by H. F. Reid (1906) is an explanation of how energy is released during an earthquake. As the earth's crust deforms the rock which span the opposing sides of a fault are subjected to shear stress. Slowly they deform when their internal rigidity is exceeded.

5.3.2 Seismic waves

Seismic waves are elastic waves generated during an earthquake. These waves travel either along or near the earth's surface or through the earth's interior. Waves travelling near the earth's surface are called *surface waves* and those travelling in the interior of the earth are called *body waves*. Body waves includes; *Primary waves* (P-waves) and *Secondary waves* (S-waves). Surface wave includes *Rayleigh* and *Love waves*. Surface waves are slowest but cause most of the surface damage.

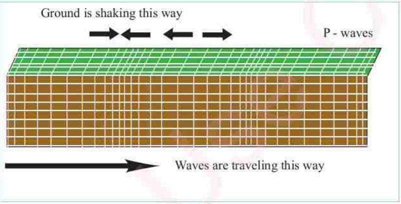


Figure 5.33 Primary waves propagation

(a) Primary waves (P-waves) or compressional waves are seismic waves that move by pushing the ground back and forward in the same direction and in the opposite direction as indicated by arrows in the direction of the waves as shown in Figure 5.33.

The primary waves, like sound waves, are longitudinal waves which can propagate through both solids and liquids. The velocity of P-waves V_p is obtained from the expression given by:

$$V_p = \sqrt{\frac{K + \frac{4}{3}\propto}{\rho}} \qquad (5.30)$$

where K is bulk modulus, $\propto$ is shear modulus and ρ is density. Primary waves travels rapidly through solids and more slowly through liquids and gases.

(b) Secondary waves (S-waves) or shear waves are seismic waves that shake the ground back and forth at right angle to the direction of wave propagation, as shown in Figure 5.34. Unlike the P-waves, the S-waves are transverse.

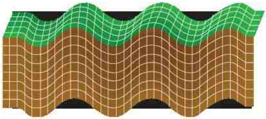


Figure 5.34 Secondary waves propagation

The velocity of secondary waves V_s is given by the expression:

$$V_s = \sqrt{\frac{\propto}{\rho}} \qquad (5.31)$$

where $\propto$ is shear modulus and ρ is density. Secondary waves travel in all directions away from their source, at speeds which depend upon the density of the rocks through which they are propagated. On the surface of the earth, S-waves are responsible for the sideway displacement of walls and fences, leaving them with 'S' shape. In fluids $\propto = 0$, hence S-waves cannot travel in fluids (liquids and gases).

(c) Rayleigh waves are seismic surface waves causing the ground to shake in an elliptical motion, with no transverse motion, as shown in Figure 5.35.

Figure 5.35 Rayleigh waves propagation

(d) Love waves are surface waves having a horizontal motion that is transverse (sideways shaking) to the direction of propagation, as shown in Figure 5.36.

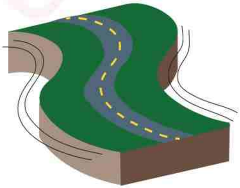


Figure 5.36 Love waves propagation

When an earthquake occurs, P-waves and S-waves travel in all directions within the interior of Earth. Surface waves travel exclusively at the surface of the earth with a slower speed than the P- and S-waves. S-waves travel more slowly than P-waves. The speed of P- and S-waves as given by equations 5.30 and 5.31 respectively, depend on the physical properties of rocks. The harder the rock, the faster the waves. Therefore, waves speed depends on rock density, pressure, temperature and mineral composition. The relationship between the speeds of seismic waves and rock density combined with the properties of wave refraction and reflection at the interface between earth's layers, can be used to reconstruct the structure of earth's interior.

5.3.3 Earthquake localization and magnitude

Seismic waves' amplitude and frequency can be measured using an instrument called a seismometer Figure 5.37. A seismometer operates on the principle of the differential motion between a free mass and a supporting frame fixed on the ground to record the seismic waves. Basically the instrument consists of a frame fixed to the ground, a heavy mass that is loosely attached to the frame and equipped with a pen, and a revolving clock-driven drum whose axis is fixed to the frame, with a special paper wrapped around the drum. When seismic waves arrive at the location the frame vibrates with the ground. Since the heavy mass is only loosely attached to the frame, it does not follow ground vibrations and consequently can record the difference in motion between the mass and the frame on the revolving drum. A mass attached to a vertical spring is used to record vertical ground movements, Figure (5. 37 (a)) whereas a mass attached to a hinge is used to detect horizontal ground movements, Figure 5.37 (b). The data are recorded as *seismograms,* (Figure 5.38) they can be used to determine the magnitude of the earthquake, direction of the seismic waves location of the epicentre, focus of the earthquake and the medium through which the waves have been travelling.

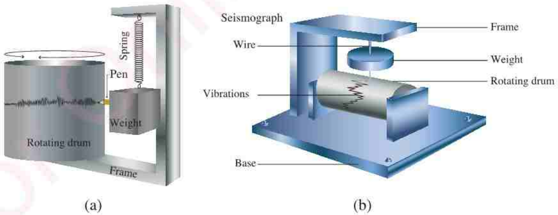


Figure 5.37 A basic seismometer to record (a) vertical ground movements and (b) horizontal ground movements.

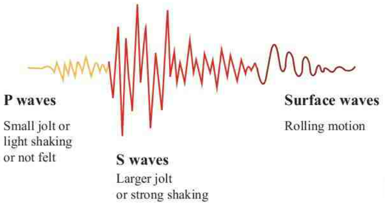


Figure 5.38: Seismograms for different types of seismic waves

5.3.4 Determination of the epicentre of an earthquake

P-waves always travel faster than S-waves, hence they are the first to be detected and recorded by a seismometer. The time difference between the arrival of the two waves is proportional to the epicentral distance i.e the distance between the epicentre and the seismograph station. P-waves and S-waves epicentral distance-time curves have been established and can be used to determine the distance between the seismograph station and the epicentre of an earthquake. If at least three seismographs in different geographical locations record the same seismic waves, precise location of the epicentre of the earthquake can be determined as described below:

(a) Determine the time difference between the arrivals of P-waves and S-waves on the seismograms obtained from seismograph recorded, say, from station A as shown in Figure 5.39. The S-P time interval is 5 minutes.

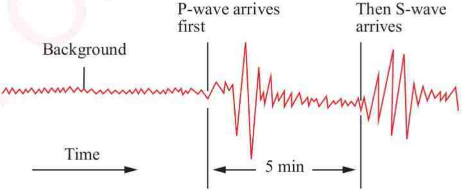


Figure 5.39: Differential time of arrivals between P- and S-waves.

(b) Take the seismic travel-time curves and identify the epicentral distance corresponding to the time difference of 5 minutes. In Figure 5.40 this is about 2300 km. Since the direction from which the waves came, is not known the location of the epicenter could be anywhere along the circle of radius 2300 km around the seismograph station.

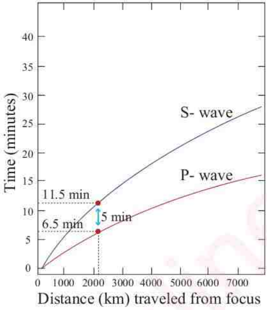


Figure 5.40 Epicentral distance from distance - time curves.

(c) Repeat procedures (a) and (b) for at least two additional seismograph stations say B and C. Assume epicentral distances of station B and C were determined to be 7000 km and 4800 km respectively. The intersection of the three circles drawn using 2300 km, 7000 km and 4800 km as the radii give the position of the epicentre as shown in Figure 5.41.

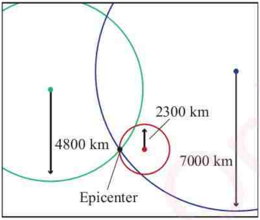


Figure 5.41 Determination of an earthquake epicenter from three stations.

5.3.5 Magnitude of an earthquake

The magnitude of an earthquake is a quantitative measure of the size of the earthquake. There are different magnitude scales which define the size of earthquakes. The most common one is the Richter scale proposed by Professor Charles Richter who defined the local magnitude M_L of an earthquake observed at a station as:

$$M_L = \log_{10} A(\Delta) - \log_{10} A_o(\Delta) \qquad (5.32)$$

where A and A_o are the largest amplitudes of the earthquake event and the standard earthquake event, respectively, and Δ is the epicentral distance. The local magnitude gives the same size of an earthquake event no matter where the recording station is located.

According to Wood Anderson seismometer with a trace amplitude 1μm and at a distance of 100 km, the local magnitude, M_L is defined such that the magnitude zero earthquake value of $\log(A_o)$ will be written as:

$$\log(A_o) = \log(A_r) - \log(A_{100}) \qquad (5.33)$$

Table 5.2 Earthquake level, magnitudes (M), occurrence frequency and effects

Level	Magnitudes	Annual occurance	Effects
Great	> 8	1	Near total destruction, massive loss of life
Major	7 – 7.9	17	Severe property damage, large loss of life
Strong	6 – 6.9	134	Property damage, loss of life
Moderate	5 – 5.9	1319	Property damage
Light	4 – 4.9	13,000 (estimated)	Some property damage
Minor	3 – 3.9	130,000 (estimated)	Felt by humans
Very minor	< 3	1,300,000 (estimated)	Hardly detected

Source *British Geological Survey*

where A_{100} is the maximum trace amplitude at 100 km from the source and A_r is the amplitude at a certain distance r. For epicentral distances different from 100 km, a correction factor has to be introduced in equation (5.33). This has led to several alternative formula, most of which take the distance (D) as a direct variable rather than a parameter in the form of:

$$\mathrm{M_L} = c_1 \log_{10} \mathrm{A} + c_2 \log_{10}(D) + c_3 \quad (5.34)$$

where c_1, c_2 and c_3 are correction factors. The most famous one is the Lillie's empirical formula:

$$\mathrm{M_L} = \log_{10} \mathrm{A} + 2.76 \log_{10}(D) - 2.48 \quad (5.35)$$

The level, magnitude, occurrences and effects of earthquakes around the earth are given in Table 5.2.

Exercise 5.3

1. (a) Why are some earthquakes stronger than others?
   (b) How can an earthquake of a moderate magnitude have a high intensity?
2. (a) Explain the formation of an earthquake using the elastic rebound theory.
   (b) What is the difference between a seismograph and a seismometer?
3. How does a seismologist use the graph of time versus distance for seismic waves to find the location of an earthquake's epicenter?
4. (a) State the consequences of earthquakes.
   (b) What precautions do engineers take to construct a building that is more likely to survive the effect of an earthquake?
5. At a recording station a difference in time of arrival between P-waves and S-waves was observed to be 1.5 s. What is the approximate distance from the station at which the event occurred? Assume P-waves and S-waves velocity to be 4 km/s and 2 km/s respectively.
6. Construction of dams often increases earthquake activity. Justify.

5.4 Environmental pollution

The environment can be defined as the natural and technological, social, and cultural world that surrounds us. Pollution is the introduction of substances into a natural environment that causes instability, disorder, harm or discomfort to the ecosystem i.e. physical systems or living organisms. Pollution can be in the form of chemical substances or energy, such as noise, heat, or light. Pollutants are waste or foreign materials or energy that pollute the environment such as air, water or soil. The severity of a pollutant is determined by its chemical nature, concentration and persistence. In this section, the types of atmospheric pollutants and their sources, transport mechanism of pollutants and their effects on the environment will be discussed. Also identification of sources of nuclear waste and methods of its disposals will be treated.

5.4.1 Atmospheric pollution

The atmosphere is a layer of gases surrounding the earth. It provides air to living organisms, shelters the earth from harmful solar and other celestial radiations; it also warms the earth's surface via the greenhouse effect. Introduction of chemicals, particulate matter, or biological materials which change the atmospheric natural environment and consequently the wellbeing of living organisms constitute air pollution.

(a) Types of air pollutants

Air pollutants can be either primary or secondary. Primary pollutants are directly emitted from either natural events or human activities while secondary pollutants are introduced into the atmosphere as a result of interactions of primary pollutants. The natural events associated with pollution include dust storms and volcanic eruption, human activities including agriculture, transportation, industrial activities and household emissions. Over 90% of global air pollution is due to five primary pollutants which are carbon oxides (monoxide and dioxide), Nitrogen oxides, Sulphur oxides, Hydrocarbons and particulate matter. An important example of a secondary pollutant is ozone which is formed near the surface of the earth when hydrocarbons and nitrogen oxides combine in the presence of sunlight. Air pollutants that need air quality standards are carbon monoxide (CO), particulate matter (PM), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), lead (Pb) and ozone (O_3).

(b) Sources, effects and mitigation of air pollutants

(i) Carbon monoxide

Carbon monoxide (CO) is a colourless, odourless, toxic and flammable gas originating from an incomplete combustion, of materials with oxygen whereas carbon dioxide (CO_2) is a complete oxidation product.

Sources of carbon monoxide: Sources of carbon monoxide include automobile exhausts, coal burning, natural gas or biomass burning. Atmospheric oxidation of methane gas and other hydrocarbons also produces carbon monoxide.

Effects of carbon monoxide: Carbon

monoxide may cause illness such as headaches, dizziness and drowsiness and if excessive gas is inhaled, heart failure or permanent brain damage.

Strategies to reduce carbon monoxide: Carbon monoxide emission can be reduced by ensuring complete combustion in vehicles and when burning fossil fuels or wood.

(ii) Sulphur dioxide

Sulphur dioxide (SO_2) is a colourless gas with a choking or suffocating odour.

Sources of sulphur dioxide: These include chemical industries and oil refineries.

Effects of sulphur dioxide: Sulphur dioxide reacts with moisture in human or animal organs to form strong irritating acid. Sulphur dioxide can combine with moisture in the atmosphere to form dilute sulphuric acid or sulphate particulates that contribute to haze in the atmosphere.

Strategies to reduce sulphur dioxide: Sulphur dioxide emissions can be mitigated in several ways including improved efficiency of fuel to electricity conversion. Also through lowering sulphur contents in fuels (eg washing the coal) before use, during combustion eg fluidized bed integrated gasification combined cycle and after combustion using different technologies such as fuel gas desulphurization.

(iii) Nitrogen dioxide

Nitrogen dioxide is a reddish brown irritating gas.

Sources of nitrogen dioxide: Sources of nitrogen dioxide include motor vehicle exhausts, fossil fuel combustion processes, volcanic eruptions and lightning.

Effects of nitrogen dioxide: Nitrogen dioxide on humans causes increased likelihood of respiratory problems.

Strategies to reduce nitrogen dioxide: Organic nitrogen content in fuels cannot be lowered before combustion, however it is possible to be reduced during combustion by using fluidized bed "combustion technique". After combustion, nitrogen oxide can be reduced by treating the emitted gases through techniques such as selective catalytic reduction, selective non-catalytic reduction and activated carbon process.

(iv) Hydrocarbons

These are volatile organic compounds that can evaporate easily

Sources of hydrocarbons: Hydrocarbon sources include organic solvents, paints, and other organic chemicals.

Effects of hydrocarbons: These gases are toxic to living organisms and contribute to ground-level ozone formation. Ground-level ozone is the result of chemical reactions between volatile organic compounds and nitrogen oxides in the presence of sunlight. However, most ozone is produced naturally in the upper atmosphere or stratosphere. Even though both types of Ozone contain the same molecule their presence in different parts of the atmosphere has very different environmental consequences. Ground-level ozone has adverse effect on human health and significant impact on vegetation including reduced productivity of some crops. The stratospheric ozone blocks harmful solar utra violet from reaching the earth's surface thereby protecting life on earth from radiation effects. The emissions of chlorofluorocarbons (CFCs) into the atmosphere depletes the ozone in the stratosphere.

Strategies to reduce hydrocarbons: Hydrocarbons can be reduced by improving the combustion of hydrocarbon based gasoline and diesel fuels. Reduction of dependency on fossil fuels is another way of reducing this pollutant to the environment.

(v) Particulate matter in the atmosphere

Particulate matter also known as aerosol particles or suspended particles is a complex mixture of both organic and inorganic particles such as pollen, dust, soot, smoke, fog, mist, fumes, and liquid droplets.

Sources of particulate matter: Primary sources of particulate matter include combustion of different types of fuels, civil works such as construction, quarrying, demolition, agricultural and mining activities. Natural sources like wind-blown dust and wild fires also contribute to particulate matter in the atmosphere. Secondary particulate matter sources directly emit air contaminants into the atmosphere that form or help form particulate matter. These pollutants include gases such as sulphur dioxide, the oxides of nitrogen, ammonia and volatile organic compounds. These gases react in the atmosphere to form solid sulphates and nitrates. Organic aerosols may also be formed by the oxidation of volatile organic compounds. Most of these secondary pollutants originate from industrial activities, vehicles especially those with diesel engines.

Effects of Particulate Matter: The main problem to human caused by atmospheric particulate matter is how far it is able to penetrate through the respiratory system leading to lung and heart diseases. *Strategies to reduce particulate matter*: Particulate matter pollution can be minimized by reducing the amount of particulate matter such as smoke produced in industries and transportation activities.

(c) Transport mechanism of atmospheric pollutants

At local level, the primary factors affecting the transport and dispersion of atmospheric pollutants are wind and stability of the atmosphere. As the particulate matter in the atmosphere are light, they are carried away by wind. The pollutant usually goes in the direction of the wind, the higher the wind speed the higher the dispersion and consequently the lower the concentration of pollutants in the immediate area. Also pollutants transport in the atmosphere is affected by the vertical motion of air in the atmosphere which is directly correlated to different types of weather systems.

Air near the surface of the earth is usually warmer during day time as the earth absorbs sun's energy. Since in the upper atmosphere the air is cooler, warmer and lighter air rises up causing unstable conditions in the atmosphere as shown in Figure 5.42. This condition results in dispersal of pollutants in the air.

Physics Form VI X.indd 251 10/11/2019 17:25

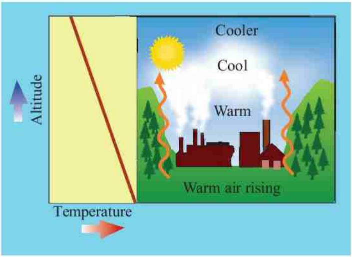


Figure 5.42 Free vertical air movement of pollutants in unstable atmosphere

In practice, due to presence of materials that absorb sunlight, warm air is above cool air as shown in Figure 5.43, the air mixing dynamics are significantly reduced. Temperature or thermal inversion in the troposphere leads to a stable atmosphere. Hence, the inversion creates a localized air pollution as it hinders the rise and dispersal of air pollutants.

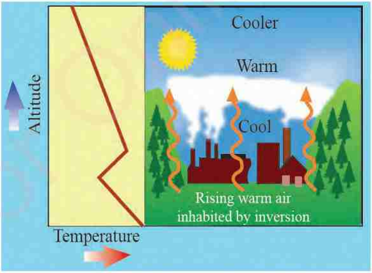


Figure 5.43 Vertical air movement through stable atmosphere

5.4.2 Nuclear waste

Nuclear waste is a radioactive material (fission fragments) produced in a nuclear reactor based on nuclear fission. Materials that constitute nuclear waste include $^{141}_{56}Ba$ and $^{92}_{36}Kr$ fragment. Nuclear waste should not be mistaken for radioactive waste as by-products of nuclear fuel processing, spent radioactive sources such as radioactive cobalt used for cancer treatment and other radioactive sources used in industry and agriculture.

Both nuclear and radioactive wastes have adverse health effects in living organisms including human beings. The severe effects commonly known as *somatic* effects are deterministic in the sense that exposure to radioactive substances at higher level produces immediate effects which are known to occur under such circumstances. The stochastic or statistical effects are due to exposure to low level of radioactivity and do not occur immediately. Cancer, is among the examples of stochastic effects which occur between 10 – 15 years after exposure. Since foetus undergoes rapid cell division, low level exposure can cause birth defects or prenatal death.

(a) Types of nuclear waste

Nuclear waste is classified as low, intermediate and high level wastes. For example every nuclear facility has cleaning tools such as mop heads and protective clothing which could be contaminated by radioactivity forming low level waste during disposal. Intermediate level waste consists of reactor parts such as water filters that have been irradiated and become radioactive. High level waste is essentially the spent nuclear fuel.

(b) Methods of disposal of nuclear waste

Disposal systems of long lived waste must be designed and constructed in such a way that they can last for thousands of years and be robust enough to prevent leakage of highly radioactive materials. Low level waste can be disposed-off by shallow land burial facilities designed, constructed and operated to meet safety standards. Intermediate and high level nuclear wastes from power plants are disposed in deep geological repository. However before disposal, the high level nuclear waste must be stored under water for five years before they go through *vitrification* process. In this process, the nuclear waste is mixed with glass-forming chemicals in a furnace. When the molten glass solidifies, it is placed in canisters which subsequently immobilize the waste, as shown in Figure 5.44

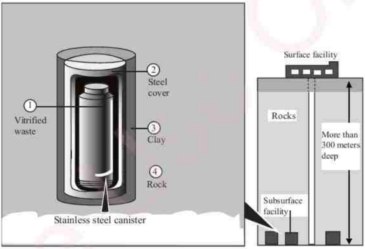


Figure 5.44 High-level radioactive waste disposal site

5.4.3 Optical properties of atmospheric particulate matter and visibility

Atmospheric pollution often affect visibility which is defined as the greatest distance at which one can see an object along the horizon. Many factors determine visibility in the atmosphere including, brightness of the sky, how good human eyes are and the atmospheric concentration of constituents especially the particulate matter. Visibility is reduced when the particulate matter between the observer and the object absorbs or scatters light.

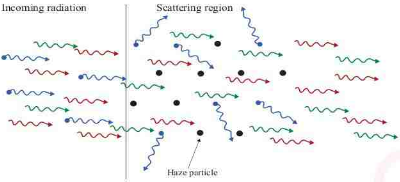


Figure 5.45 Scattering of blue light when a beam of white light passes through a haze made up of small particles.

The size, concentration and chemical nature of the particles, all of which affect optical properties of the particulate matter, determine visibility of the lower atmosphere. Light scattering is a general physical process whereby a light beam is forced to split and travel in direction different from the initial direction. This is caused by the interaction of the light with atmospheric. The magnitude of scattering depends on the particle size and the wavelength of the incident light falling on the particulate matter such as smog, haze, smoke, dust and snow. If the particle is assumed to have a dimension d_p and the wavelength of incident light is λ, then the following can be deduced:

(a) If $d_p << \lambda$ light is not scattered to great extent and hence light rays go directly to the observers eyes.

(b) If $d_p >> \lambda$ significant light scattering occurs as for case of haze whose dimension is closer to the wavelength of blue light as shown in Figure 5.45.

(c) If $d_p >> \lambda$ the light is strongly absorbed or reflected.

Exercise 5.3

1. The pollutants of primary concern are the criterial pollutants which include carbon monoxide (CO), nitrogen oxides (NO_x), sulphur oxides (SO_x), particulate matter, lead (Pb) and ozone (O_3). For each of pollutants explain its:
   (a) origin;
   (b) effects on health and environment; and
   (c) mitigation methods.
2. Explain atmospheric temperature inversion. Describe how this inversion affects the dispersion of air pollutants in the atmosphere.
3. Radioactive isotopes are used in hospitals, in industry, and in agriculture. The concern has always been on how to dispose them after use.
   (a) Using examples, explain the differences in properties between low-level and high-level radioactive waste.
   (b) Explain how human activity generates radioactive wastes.

(c) Identify one adverse effect on human health that can result from exposure to radioactive wastes.

4. How is nuclear waste produced and disposed
5. Explain why air pollution is a much larger problem in urban areas than in rural areas.
6. Why carbon dioxide (a non-toxic normal component of the atmosphere) is called a "greenhouse gas"?
7. One of the major environmental concerns, is the destruction of the ozone layer. What are the reasons for this concern?

Revision exercise

1. Explain the influence of latitude and seasons on the unequal heating of the atmosphere.
2. Explain why the temperature in the troposphere decreases with an increase in height?
3. Why is solar energy considered as the source of most energy types on earth?
4. Describe the three components of solar radiation and discuss their uses or effects.
5. Discuss how the greenhouse effect arises.
6. How does soil particle size affect texture and water holding capacity of the soil?
7. Apart from crop growth, what are other possible uses of the soil?
8. Distinguish among sand, silt, clay, and loam soil.
9. What factors influence the flow of water through soils?
10. What constitutes a mulch?
11. What climatic factors influence the evaporation of water from the soil?
12. (a) State the Darcy's law as applied to flow of water in soil.
    (b) State limitation of Darcy's laws.
    (c) How is Darcy's law related to Ohms law?
13. What drives water flow in the soil?
14. Describe the conceptual model of water flow in soil.
15. Poiseulle's law is a special case of Darcy's law. What term in Poiseulle's law, corresponds to K in Darcy's law?
16. Describe an experiment to measure the coefficient of soil permeability or hydraulic conductivity using the following methods:
    (a) constant head test;
    (b) falling head test; and
    (c) pump out test.
17. Figure 5.46 represents a steady state one dimensional seepage situation in which the upstream and downstream water levels are maintained constant. The container in which the soil with permeability 1.2×10^{-5} m/s and porosity $n = 0.2$ is placed consists of two sections having different diameters. The areas of the upper and lower sections are 28 cm^2 and

14 cm^2 respectively. Determine:

(a) the rate of seepage flow through the soil; and

(b) the actual velocity of flow through the upper section of the soil.

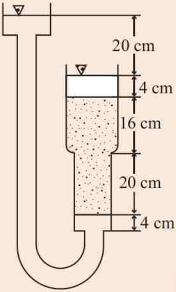


Figure 5.46 One dimensional steady state

18. Why is soil temperature more important than air temperature for plants?
19. Describe the mechanisms of heat transport in soil by conduction.
20. What is meant by the following terms as applied to heat transfer in soil:
    (a) specific heat;
    (b) thermal conductivity;
    (c) thermal diffusion;
    (d) thermal resistance; and
    (e) thermal admittance.
21. State the factors on which soil specific heat, thermal conductivity and soil thermal diffusion depend.
22. Determine the direction and quantity of heat flow per unit time that will flow in one day, when the dry loam soil temperature at the surface is 33 ϒC and at 3.67 m depth it is 24 ϒC. Assume the thermal conductivity of a dry loam soil to be $0.15\ \text{W/mK}$.
23. You are provided with a soil column containing 32 cm of dry sand over 12 cm of dry loam soil. The ends of dry sand and dry loam soils are attached to a constant temperature bath with the top maintained at 22 °C and the bottom at 5 °C. Computed:
    (a) the steady state heat flux through the two layers; and
    (b) the temperature at the sand-loam interface.

    Use:
    - Thermal conductivity of dry sand soil $= 0.27\ \text{W/mK}$
    - Thermal conductivity of dry loam soil $= 0.15\ \text{W/mK}$
24. What does the term photovoltaic mean?
25. (a) Define renewable energy
    (b) Explain why solar energy is considered to be renewable energy.
26. An old photovoltaic module of area 0.8 m^2 has the following parameters: $I_{sc} = 5.2\ \text{A}, V_{oc} = 21.7\ \text{V}$, $V_{mp} = 15.6\ \text{V}$, $I_{mp} = 4.7\ A$ recorded at an irradiance of $1000\ \text{Wm}^{-2}$.
    (a) Calculate the fill factor, (FF).
    (b) Calculate the efficiency of the module.

27. Given the I-V curve in Figure 5.47
    (a) Estimate the current and voltage at maximum power point.
    (b) Determine I_{sc}, V_{oc} and FF.

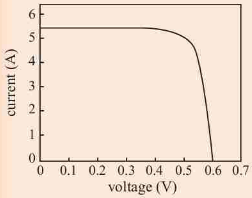


Figure 5.47 Current –Voltage curve

28. What are the environment benefits of geothermal energy?
29. What is a geothermal heat pump?
30. What do you understand by wind power?
31. Explain physical processes in the earth's atmosphere that lead to wind formation.
32. The Tip Speed Ratio (TSR) is an extremely important factor in wind turbine design. Explain
33. Describe the working mechanisms of two types of wind turbines blades.
34. Explain how electricity can be generated from wind energy.
35. If the wind speed is 11.5 m/s and the speed after the turbine is 8 m/s, what is the power extraction coefficient of this wind turbine? (Air density 1.225 kg/m^3)
36. Explain the effect of wave and tidal energy converters on the ecology of the area, on shipping and navigation.
37. Show that the wave power (P) in a given water wave of height H is given by:

$$P = \frac{1}{32\pi}\rho g^2 TH^2$$

where ρ the density of seawater, T is the period and g the acceleration due to gravity.

38. Explain how the position of the sun and the moon relative to the earth's motion affect the tidal energy resource.
39. Describe how tidal energy and wave energy are converted into electricity.
    If there are two high tides around the earth at any instant, explain why the interval between successive high tides is 12 hours 42 minutes rather than 12 hours.
41. Assuming the speed v of surface waves on deep water depends only on the acceleration due to gravity g and the wavelength λ, use dimensional analysis to derive an algebraic expression of the form $v = kg^a\lambda^b$, where k is a dimensionless constant. Calculate the speed of a surface wave on deep water of wavelength 100 m.
42. (a) What are the causes of earthquakes?

(b) Where do most earthquakes occur ?

43. The epicentral intensity of an earthquake that occurred in 1870 is estimated to be IX in the Modified Mercalli Intensity (MMI) scale. Estimate the approximate magnitude of the earthquake.

44. (a) What are seismic waves?

(b) Explain the terms elastic rebound and seismology

45. Explain briefly the ozone layer depletion.

46. Describe three characteristics of an ideal deep underground storage site for high-level radioactive waste.

47. Do you think nuclear power should be used instead of natural gas for generating electricity in Tanzania? Explain your views against and in favour of each energy source?

Chapter One

Exercise 1.1

2. $2\times10^{-7}\,\Omega\,\text{m}$
3. (a) 2A (b) 4A
4. $J = 2.5x10^{6}\,\text{Am}^{-2}$ and $v_d = 1.83x10^{-4}\,\text{m/s}$
5. Aluminium
6. (a) 6 Ω (b) 6 Ω
7. (a) 2.73 Ω (b) 2.83 A
8. L_{Nichrome} = 1.5 m, 1.51
9. (a) (i) 17.2 ^{0}C
   (ii) $1.2\times10^{-7}\,\Omega$ m
10. (a) 0.012 A, 0.870 A, 0.859 A
    (b) 11.915 V

Exercise 1.3

1. (a) 6.28 Ω (b) 31.42 Ω (c) 628 kΩ
3. (a) 0.503 A (b) 282.84 V
4. 4 mΩ
5. 3.22 Ω, 15.5 A
6. 17.36 A, zero
7. 5.43 A, zero
8. 2,000 W
9. (a) 1.29 *A* (b) 3.19 ms
10. 81.6 rad/s, 27.2, 3 rad/s

Revision Exercise

2. 16 C, 0.27 A, 33.75 Ω
3. 1.8×10^{25} electrons, $1.56\times10^{-4}\,\text{m/s}$
4. $\left(3/4\right)R$, 0.72 Ω
5. 0.02 $^oC^{-1}$
6. 0.004 $^\circ$C^{-1}, 50 $^\circ$C
7. (a) 8.16 m, (b) 2.04 m
8. (b) (i) 1.5 V (ii) 4 Ω
   (iii) 1 Ω (iv) 0.75
10. 6 A
11. $Y = 2000\,\Omega$ and $R_{voltmeter} = 1000\,\Omega$
12. 0.83 Ω, 12 A
13. 1.5 A, 2 A, 0.5 A, 2.5 A, 0 A
14. $v_1 = 22$V, 5V, 1 Ω
15. (a) $R_1 = 1.2\,\Omega$, V_1=18.3 V, V_2=13.17 V
    (b) V_1=18.3 V, V_2=13.17 V
16. (d) (i) 1.53 V, (ii) 60 cm
17. (a) $L_{MN} = 82.5\,\text{cm}$ (b) $V_{reading} = 1.3\text{V}$
18. $I_g = 0.06$ A
19. 1.69 Ω
20. 50 Ω
26. (a) 6.38 mA (b) 10.16 V (c) 57.860
    (d) 0.0766 W
27. 19.89 Hz, Q = 25
28. 62.4^0, 0.43
29. (a) 40 Ω (b) L=1.18 H (c) 83.9^0

Chapter Two

Exercise 2.1

1. $3\times10^{-7}\,\text{T}$
2. $B_x = -2\text{T}$
3. 5.29 m
4. (a) 0.26 m (b) 2×10^{6} Hz
   (c) 2500 eV

5. 0.392 A, direction of current is to the right
6. 4.16 N
7. 3.48×10^{-26} Nm
8. 0.024 Am^2
9. (a) 6.52×10^{-8} s
   (b) 1.11 x 10^{-11} J

Exercise 2.2

1. (a) 0.013 T (b) 7.85 mT
2. 1.58 x 10^{-3} T
3. 6.28×10^{-5}T
4. (a) 0.002 N

Exercise 2.3

3. 0.4 H
4. 1350 turns
6. 64 J
8. 0.219 A
9. (a) $L = 2.30\times10^{-3}$ H
   (b) $\varepsilon = 0.023$ V
10. (a) 0.377 H (b) 188.5 V

Exercise 2.4

6. 1.13 T
7. (b) 2.1×10^{-4} Am^2 (b) 2.7

Revision exercise

4. (a) 5.6×10^5 V/m (b) 8400 V
5. (c) 5×10^{-12} N (d) 5.7 T
   (e)(i) 1.65×10^5 V/m
   (ii) $2.48\times10^4 V$
6. 43.77Nm
8. (a) 31.83 A (b) 0.4Nm (c) 0.283 Nm
13. 0.78 T
16. 2.5×10^{-23} kg
24. 4.54×10^{-4}Wb
25. 0.0197 T
27. 3.14×10^{-5} V
28. 2B
34. (a) 0.05 V (b) 0.0027 A (c) 0.13 mW
37. 15μ C
46. 1×10^{-5}
54. 3456 V
64. 1389

Chapter Three

Exercise 3.2

9. 19 V, 7 mA, 2.14 k Ω

Exercise 3.3

5. $R_L = 3.23$ kΩ, $R_L = 310$ kΩ
6. (a) I_C =6.47 mA, I_B=0.129 mA,
   (b) I_E=3.1 mA, α = 0.968, β = 30.25

Exercise 3.5

7. 1.05 V
8. (i) $-0.72\cos 40t$, (ii) -1.62×10^{-2} V

Exercise 3.6

7. (a) 6.2 kHz (b) 197 kHz
   (c) 2031.1 kHz
8. M=0.33

Revision Exercise

10. (a) 100 (b) 1000Ω (c) 500
11. (i) 2.24m A, (ii) 10.16 V
19. $9526\,M\Omega,\ 0.0157\Omega$
22. (a) M= 0.5
    (b) upper side band = 1010 kHz, and lower side band = 990 kHz

Chapter Four

Exercise 4.1

3. $m_p = 2.0143u$, the particle is ${}^{2}_{1}\mathrm{H}$
4. 5 V
5. 196.8 nm
6. X, because its work function is less than photon energy $(\phi < hf)$
8. $\lambda = 282\,\mathrm{nm}$, is low than 400 nm (visible range)
9. Increase current
10. 7.07 eV

Exercise 4.2

3. (a) $(a)\,n = 4\ (b) = 6.1\,\mathrm{nm}$
   (b) 91.2 nm
4. $L_a = 656.1$ nm and $L_b = 486.3\,\mathrm{nm}$
5. 122.4 eV and 10.1 nm

Exercise 4.4

2. (a) (i) $2.33\times10^{-3}\mathrm{u}$
   (ii) $3.023\times10^{-2}\mathrm{u}$
   (b) $2.17\,\mathrm{MeV}\left({}^{2}\mathrm{H}\right)$ and $28.1\,\mathrm{MeV}\left({}^{4}_{2}\mathrm{H}\right)$
4. 36.12 GBq

6. 7.27 MeV
7. $Q_1 = 2.23\,\mathrm{MeV}$
   $Q_2 = 23.85\,\mathrm{MeV}$
   $Q_3 = 3.25\,\mathrm{MeV}$
8. 8 alpha particles and 6 beta particles
9. 235 MeV
10. 1039.62 MeV

Revision Exercise

6. 2.91×10^{30} photons/s
7. 4.13 eV
8. 0.69 eV
10. (a) 4.13×10^{-2} V
    (b) $1.033\,\mathrm{eV}$ or $1.65\times10^{-19}\,\mathrm{J}$
11. (a) 2.21 eV
    (b) $9.84\times10^{5}\,\mathrm{m/s}$
    (c) 2.52×10^{18} photons $\mathrm{m^2s}$
12. (a) 2.8 eV
    (b) 0.50 eV
    (c) 539 nm
13. $V_S = 1.85\,\mathrm{V}$
22. (a) $0.145\,\mathrm{nm}$
    (b) $4.34\times10^{-13}\,\mathrm{m}$ *or* $0.434\,\mathrm{pm}$
24. $9.05\times10^{-4}\,\mathrm{nm}$
32. (a) 48.36 volts
    (b) 54.4 eV
33. $n = 3$
34. 7.65 eV
37. $(a)\,3.10\mathrm{x}10^{-19}\mathrm{J}$
    $(b)\,0.155\mathrm{J}$
    $(c)\,4.2\mathrm{x}10^{16}$ photos
39. Mass number $A = 7$

40. (a) 1.00783 u
(b) 6.82 MeV
(c) 4.95 MeV

46. (a) $7.07x10^7$ Atoms
(b) 10564 years

48. 2.5 hours

50. 1.57×10^9 kg/s

55. 3.608×10^{16} fissions/s

Chapter Five

Exercise 5.1

6. (a) 0.103 cm/ hr (b) -19.39 cm

8. (a) $49.5\ \text{Wm}^{-2}$ (b) $1.78 \times 10^5\ \text{Jm}^{-2}$

Exercise 5.2

9. (a) $V_{oc} = 0.61\text{V}$
(b) $J_{mp} = 192.98\text{A}$
(c) $\dfrac{P_{mp}}{A} = 105.95\text{Wm}^{-2}$
(d) $\eta_{\max} = 11.15\%$
(e) 0.425m^{-2}

13. 2.82 m

Exercise 5.3

5. 6 km

Revision exercise

15. $$K = \frac{\rho g r^2}{8\eta}$$

17. (a) $1.2 \times 10^{-2}\text{cm}^3/\text{s}$
(b) $2.14 \times 10^{-3}\text{cm/s}$

22. $3.24 \times 10^5\ \text{Jm}^{-2}$

23. (a) $8.5\ \text{Wm}^{-2}$
(b) $11.9\ ^\circ\text{C}$

26. (a) 0.65
(b) 9.1 %

27. (a) $I_{mp} = 5.3\text{A}$ and $V_{mp} = 0.5\text{V}$
(b) $I_{sc} = 5.4\text{A}$, $V_{oc} = 0.6\text{V}$ and $FF = 0.85$

35. $C_p = 0.44$

43. An earthquake of magnitude 7

Table of physical constants

Name	Symbol	Constant
Permeability of free space	$\propto_0$	$4\pi \times 10^{-7}\, TA^{-1}\text{m}$
Permittivity of free space	ε_0	$8.854 \times 10^{-12}\,\text{Fm}^{-1}$
Radius of the Earth	r	$6.4 \times 10^{6}\,\text{m}$
Electronic charge	e	$-1.602 \times 10^{-19}\,\text{C}$
Planck's constant	h	$6.626 \times 10^{-34}\,\text{Js}$
		$4.135 \times 10^{-13}\,\text{eVs}$
Boltzmann's constant	k_B	$1.381 \times 10^{-23}\,\text{JK}^{-1}$
Stefan-Boltzmann constant	σ	$5.671 \times 10^{-8}\,\text{Wm}^{-2}\text{K}^{-4}$
Speed of light in vacuum	c	$2.9981 \times 10^{8}\,\text{ms}^{-1}$
Rydberg constant	R_H	$1.097 \times 10^{7}\,\text{m}^{-1}$
Mass of electron	m_e	$9.1 \times 10^{-31}\,\text{kg}$
		$0.00055\,\text{u}$
Mass of proton	m_p	$1.67 \times 10^{-27}\,\text{kg}$
		$1.00728\,\text{u}$
Mass of neutron	m_n	$1.00728\,\text{u}$
		$1.00866\,\text{u}$
Mass of hydrogen atom	m_H	$1.674 \times 10^{-27}\,\text{kg}$
		$1.0078\,\text{u}$
Atomic mass unit	u	$1.607 \times 10^{-27}\,\text{kg}$
		931.5 MeV
Mass of Helium atom		$4.00151\,\text{u}$
Mass of Uranium		$235.043929\,\text{u}$
Reduced Planck's constant	$\hbar$	$1.054 \times 10^{-34}\,\text{Js}$

Name	Symbol	Constant
		6.582×10^{-16} eVs
Fine structure constants	hc	1.986×10^{-25} Jm
		1239.853 MeVfm
	$\hbar c$	3.162×10^{-26} Jm
		197.329 MeVfm
	$\alpha=\dfrac{e^2}{4\pi\varepsilon_0 hc}$	$(137.04)^{-1}$
Bohr radius	a_0	5.292×10^{-11} m
Acceleration due to gravity	g	9.8 ms^{-2}
Gravitational constant	G	6.673×10^{-11} Nm^2kg^{-2}
Density of fresh water	ρ_w	1000 kgm^{-3}
Density of air	ρ_a	1.225 kgm^{-3}
Density of sea water	ρ_s	1024 to 1028 kgm^{-3}

Glossary

Absorption spectrum	A frequency distribution of missing intensity of electromagnetic radiation through medium due to absorption
Alternating current (*a.c.*)	An electric current which periodically reverses direction, in contrast to direct current (*d.c.*) which flows only in one direction
Amplifier	An electronic circuit device that amplifies the voltage, current, or power of a signal
Amplitude modulation (AM)	A modulation technique used for transmitting information through a radio carrier wave, whereby the amplitude of the carrier wave is modified by the signal wave
Angular momentum	A vector quantity which measures the rotational momentum of a rotating body whose magnitude is equal to the product of the angular velocity of the body and its moment of inertia with respect to the rotation axis, and pointing in the direction of the the axis of rotation
Atmospheric pollution	The introduction of harmful or excessive quantities of substances including gases, particles, and biological molecules into the Earth's atmosphere
Attenuators	Electrical devices that reduce the amplitude of a signal passing through the component, without distorting its waveform
Bandwidth	A range within a band of wavelength, frequency or energy occupied by a carrier wave
Biasing	A process of applying a *d.c.* voltage to an electronic device such as a transistor in order to establish a reference level necessary for the device to operate
Breakdown voltage	The minimum voltage that causes an insulator to become electrically conductive
Capacitive reactance	The internal impedance of a capacitor, which changes inversely with respect to frequency
Chain reaction	This is a process in which neutrons released in fission fragments produce additional fission in at least one further nucleus. This nucleus in turn produces neutrons which continue to produce fission in a repeated process.
Conductance	Numerically equal to the current flowing through a conductor per unit potential difference, or it is the reciprocal of the resistance
Conduction band	The band of electron orbitals where electrons in a solid can be exited into, from the valence band. Electrons in this band are free from binding forces of atoms and can participate in conduction of electricity

Conductors	Materials or substances that allow electric charges to drift through it, constituting an electric current
Continous spectrum	The gas pressures are higher, so that lines are broadened by collisions between the atoms until they are Smeared into a continuous. Blackbody emission is an example of continuous emission.
Current Density (J)	At point in a conductor is the electric current per unit cross-sectional area at that point
Decay constant	A proportionality constant between the size of a population of radioactive atoms and the rate at which the population decreases because of radioactive decay
Diamagnetic materials	Materials that give rise to magnetization opposite to a magnetic field that is applied to the material
Direct current (*d.c.*)	Directional flow or movement of electric charge carriers which are usually electrons
Dopants	Impurity deliberately added to a semiconductor for the purpose of modifying its electrical conductivity
Drift velocity	The average velocity of the drifting electrons acquired in a material in response to an applied electric field
Earthquake	Shaking of the Earth surface, resulting from the sudden release of stored strain energy in the Earth's lithosphere in the form of seismic waves
Electric discharge	Release and transmission of electricity in an applied electric field through a gas
Electrical conductivity	Easiness with which a material will allow an electric charge to pass through it. Mathematically, it is a reciprocal of the resistivity
Electromagnetic induction	Process by which a magnetic field induces an electrical current in the conductor moving through the field
Emission spectrum	The spectrum of radiation emitted by a substance that has absorbed energy
Energy band	A large number of closely spaced energy levels in a solid separated from other bands by energy gap
Epicenter	The point on the Earth's surface directly above the point where an earthquake starts, and where the earthquake is felt most strongly
Extrinsic semiconductor	Semiconductor doped by impurity atoms to a level that significantly alters its electrical properties
Fermi level	The highest energy level occupied by electrons at absolute zero temperature

Ferromagnetic materials	Substances which exhibit strong magnetism in the same direction of the field, when a magnetic field is applied to them
Fluorescence	The emission of electromagnetic radiation, usually visible light, caused by excitation of atoms in a material
Frequency modulation (FM)	A modulation technique that includes varying the frequency of the carrier wave on which useful information is imposed upon
Geomagnetic field	The magnetic field that surrounds the Earth, which attributed to the combined effects of the planetary rotation and the body system movement
Geothermal Energy	Heat energy of the Earth's interior which can be captured and harnessed for home and other uses
Half-life	The time taken for a given amount/activity of the radioactive substance to decay to half of its original quantity/activity
Impedance (*Z*)	Total opposition offered to the flow of alternating current
Inductive reactance	The opposition offered to electric current by an ideal inductor, when an *a.c.* voltage is applied across it
Inductor	Passive electrical component that stores energy in the form of magnetic energy. It is simply a coil of a wire
Insulator	A material with a large band gap, such that virtually no electrons are in the conduction band at room temperature, consequently no current can flow through it when subjected to an electric field
Intrinsic semiconductor	A pure form of the semiconductor without any addition of impurity atoms
Ionization	The process by which an atom or a molecule acquires a negative or positive charge by gaining or losing electrons to form an ion, often in conjunction with other chemical changes
Ionized gases	These are gases in which electrons have been accelerated and detached from the parent atoms, leaving them with net positive charge
Isotopes	Atoms that have the same atomic numbers but different mass numbers
Laser	Acronym for «light amplification by stimulated emission of radiation" it also refers to a device that produces light through a process of optical amplification based on the stimulated emission of electromagnetic radiation
Logic gates	Physical device that is used to implement Boolean function
Magnetic field	A vector field around a magnet or a loop of a wire carrying an electric current

Magnetic flux density	The force acting per unit current per unit length on a wire placed at right angles to a magnetic field
Magnetic flux	A measurement of the total magnetic lines which pass through a given area
Magnetic force	Attractive or repulsive force that is exerted by unlike poles of magnets or electrically charged particles
Magnetic intensity	Force that a magnetic field experiences on a theoretical unit magnetic pole in free space
Magnetic permeability	The tendency of material to allow the magnetic lines of force to pass through it
Magnetic susceptibility	Quantitative measure of the extent to which a material may be magnetized in relation to a given applied magnetic field
Magnetization	The process of making a substance temporarily or permanently magnet by inserting the substance in a magnetic field
Magnetosphere	A region of space in which charged particles are affected by magnetic field
Mass defect	The difference between the sum of masses of the nucleons and that of the nucleus
Meta-stable state	An excited state of an atom or other system with a relatively longer lifetime than the other excited states
Modulation Index (M)	The ratio by which the modulated and unmodulated signals vary expressed as a percentage
Modulation	Process through which information signal is superimposed on the carrier signal to be transmitted over a telecommunication or electronic medium
Monochromatic light	The light of a single wavelength or frequency
Mulching	The mechanism of covering the soil with materials such as leaves, grass clippings, shredded trees or plant for the purpose of preserving moisture and reducing weed growth
Nuclear binding energy	The minimum energy required to disintegrate the nucleus completely into its constituent nucleons
Nuclear fission	The process in which the nucleus of an atom splits into two smaller nuclei called fission fragments and the release of energy
Nuclear fusion	The process by which lighter nuclei are fused together to form a heavier nucleus with a release of energy
Operational Amplifiers	A high gain voltage amplifying device with a differential input
Paleomagnetism	The study of magnetic rocks and sediments to determine the history of the magnetic field
Paramagnetic material	Materials that are weakly attracted to a magnetic field
Phase modulation (PM)	Process of impressing information signal into a carrier wave by varying its instantaneous phase

Photocell	Light sensitive device that can be used as a sensor to detect the presence of light
Photoelectric effect	The emission of electrons from matter upon the absorption of electromagnetic radiation, such as ultraviolet radiation or x-rays
Photovoltaics	System of converting sunlight directly into electricity using semiconducting materials
p-n junction	A boundary or interface between two types of semiconductor materials, p-type and n-type
Polar wandering	The migration over the surface of the Earth of the magnetic poles of the Earth through geological time
Potentiometer	Device for accurate measurement of e.m.f. of a cell or potential difference between two points of an electric circuit
Power factor	The ratio of the real working power that is used to do work and the apparent power that is supplied to the circuit
Radioactive waste	Radioactive elements that no longer have practical uses
Radioactivity	The spontaneous emission of particles or radiation from radioactive or unstable nuclei
Radioisotopes	Atoms of the same element having same charge number but differ in mass numbers and they undergo radioactive decay
Rectification	The conversion of alternating current to direct current by using a rectifier such as a set of diodes that allow current to flow in one direction only
Repeaters	A network device that retransmits a received signal with more power and to an extended geographical region
Resistance (R)	An opposition offered by a conductor when charge flows through it
Resistivity	The opposition offered by conductor of unit cross sectional area (A) when a current flows through a unit length of the conductor
Reverse bias	The situation in which a p-n junction diode blocks the electric current in the presence of applied voltage
Seismic wave	A wave that travels through and along the surface of the earth, as result of a tectonic motion from e.g. earthquake and sometimes from an explosion
Seismometer	Instrument that measures ground motion caused by earthquakes, volcanic eruption and explosion
Seismographs	Instruments used to record details of the motion of the ground during an earthquake to determine the location and magnitude of earthquakes
Seismograms	Graph outputs by seismographs as records of ground motion at measuring station as a function of the time

Semiconductors	Materials with a band gap which is not very small or very large as compared to conductors and insulators, such that at temperatures above absolute zero, thermal energy excites electrons to the conduction band
Soil structure	The arrangement of soil separates into shapes based on its physical and chemical properties
Soil texture	The composition of the soil in terms of the amounts of small (clays), medium (silts), and large (sands) size particles that make up the soil
Solar Energy	Radiant energy emitted by the surface of the sun as a blackbody at about 5800K
Solenoid	A coil which produces magnetic field when electric current passes through it
Telecommunication	The exchange of information such as signs, messages, words, text, images and sounds by wire, radio, optical electronic and electrical means over a significant distance
Thermal runway	A self-destruction of an unstabilized transistor as a result of increased temperature
Tidal energy	Potential energy created by the surge of ocean waters during the rise and fall of tides due to gravitational attraction of the earth by the moon and sun
Transistor	A semiconductor device used to amplify or switch electrical signals.
Valence band	The outermost electron orbital of an atom of any specific material that electrons actually occupy at absolute zero temperature
Voltage divider	A circuit which provide the output voltage as a fraction of the supply
Wave Energy	The energy carried by ocean waves as they progress from one point to another
Wave-particle duality	The phenomena in quantum mechanics indicating photons and sub atomic particles can be described in terms of particles or waves
Wind energy	Renewable energy based on the kinetic energy of moving air and it can be harnessed by using a wind turbine
Windbreaks	Barrier that affect/reduces the speed of the wind. For agricultural purpose the barrier is usually made up of one or more rows of trees or shrubs to prevent breakage of plants from the effect of wind
Work function	The energy required to overcome the attractive force of an electron by the nucleus of a solids metals
Zener diode	A semiconductor device that is designed to operate in the reverse bias breakdown current levels without being damage

Bibliography

Arnold, B., & Bloor, C. (2004). *Advanced level physics.* London: John Murray Publishers Ltd.

Boeker, E., & Grondelle, R. (1998). *Environmental physics (2nd ed.).* New York: John Wiley & Sons, Ltd.

Borgaro, I., G. Ganale., & Keller, F. (1993). *Physics: classical and modern (2nd ed.).* New York, NY: MacGraw-Hill, Inc.

Duncan, T. (2000). *Advanced physics* (5th ed.). London: John Murray (Publishers) Ltd.

Illingworth, V. (1991). *The penguin dictionary of physics (*2nd ed.). London: Penguin books Ltd.

Jain, M. (1997). *Objective physics for IIT-JEE AIEEE AIIMS AIPMT and all other engineering and medical entrance examinations* (1st ed.). New Delhi: S. Chand and Company Ltd.

James, A.R, Francis, W.S., Russell, M.W., & Mark, W.Z. (1964). *Modern university physics* (1st ed.). London: Addison-Wesley publishing company Ltd.

Keith, J., Simmone, H., Sue, H., & John, M. (2000). *Advanced physics for you* (1st ed.). London: Nelson Thornes Ltd.

Keit, G. (1990). *Advanced physics (*2nd ed.). England: Cambridge University Press.

King, A.R., & Regev, O. (1997). *Physics with answers.* New York NY. Cambridge University press.

Mehta, R., & Mehta, V.K. (2009). *Principles of physics for class XII.* New Delhi: S.Chand & Company Ltd.

Mehta, K., & Mehta, R. (2009). *Principle of electronics.* Multicolour revised edition. New Delhi: S. Chand & Company Ltd.

Mehta, R., & Mehta, V.K. (2009). *Principles of physics for class XII*. New Delhi: S. Chand & Company Ltd.

Ministry of Education and Vocational Training. (2009). *Physics syllabus for advanced secondary education*. Dar-es-Salaam: Tanzania Institute of Education.

Muncaster, R. (1995). *A-Level physics* (4th ed.). Gloucestershire: Stanley Thornes (publishers) Ltd.

Nelkon, M., & Parker, P. (1995). *Advanced level physics* (7th ed.). London: Heinemann educational books.

Raju, R., & Arora, B.L. (2009). *Academic's dictionary of physics* (20th ed.). S.Chand & Company Ltd.

Robert, H. (1992). *Physics* (2nd ed.). Canada: Thomas Nelson and Sons Ltd.

Yap, E.K., & Khoo, G.K. (2008). *Essential physics SPM*. Bangi: Pearson Longman.

Index

A

absorption spectra 25
accelerated particles 23
acceptors 115
ac circuits 40
aerial environment 218
aerofoil 234
agricultural physics 215
air pollutants 249
air temperature 218
albedo 217
alternating current (a.c.) 28-40, 43, 46-50, 52, 57, 121, 122, 124, 134, 135, 139, 149, 171, 266, 268,
ambient air temperature 218
ammeters 33
ammeter-voltmeter 13
ampere 1
Ampere's law 77
amplifier 24
amplitude 44
André ampère 76
angle of dip 101
angular frequency 46
angular location 26
angular speed 62
applications of laser 197
associative law 143
atmospheric pollution 249
atmospheric window region 217
atomic mass unit 199
atomic number 198
atomic spectrum 26
attenuators 239
audio signal 139
avalanche breakdown 118
average power 35

B

back e.m.f. 35
Balmer series 191
band theory of solids 109
bandwidth 46
base 126
Becquerel 203
Betz' Law 234
biasing 134
biasing of p-n junction 117
binding energy 199
Biot-Savart law 73
black body 215
blackbody radiation 174
Blocky 221
Bohr atomic model 188
Bohr postulates 189
Boltzmann's constant 230
Boolean algebra 142
braking radiation 184
breakdown potential 21
breakdown voltage 118
Bremsstrahlun 184
bypass capacitor 135

C

capacitive circuit 37
capacitive reactance 38
capacitor 37
capacitor resistor (C-R) circuit 49
carbon monoxide 249
cell voltage 230
centripetal force 61, 62, 63, 69, 189
chain reaction 207
charged particle 62
co-axial coils 87
coefficient of self-induction 88
coercive force 98
coercivity. 98
coherent, 196
collector 126
collector-emitter 132
collimated 196
columnar 221
combination of cells 9
common base 127
common collector 127
common emitter 127
commutative law 143
comparison of the e.m.f. of cells 16
compass needles 64
complement 142
conduction band 110
conduction of electricity in gases 21
conductivity 5
conductor 1, 2, 3, 4, 5, 6, 16, 53, 58, 59, 64, 73, 75, 76, 77, 78, 79, 81, 83, 94, 104, 106, 107, 108, 111, 112, 113, 265, 266, 268
continuous spectrum 25, 184
Coulombs 1
coupling capacitor 134
critical mass 208
Crookes' dark space 22
cross-sectional area 2, 3, 4, 16, 19, 20, 53, 65, 87, 224
Curie's law 96
Curie temperature 97
current density 3
current gain 129
current loop 66
current sensitivity 68
cut off 136
cyclotron 70

D

direct current (d.c.) 24,28, 29, 31, 33,43, 55, 57 , 72, 121- 124, 134, 135,137-139, 149-152, 265, 266, 268, 272
Darcy's law 224
de Broglie wavelength 182
decay constant 203
de Morgan's theorems 144
density of copper 19
depletion layer 116
dew point 218
diamagnetic 92
diamagnetic materials 95
diffraction grating 26
diffuse solar radiation 216
diode as rectifier 121
dipole moment 67
discharged lamps 27
discharged tube 21, 24
distributive law 143
DNA 205
dopants 114
doping 114
drift velocity 2
dynamo effect 100

E

earthquake 245
earthquakes 242
effective e.m.f. 9
effective emissivity 216
elastic rebound theory 242
electrical networks 7
electric circuit 7
electric conduction 2
electric conduction in gases 21
electric field 2
electric intensity 58
electric potential energy 18
electrochemical potential 112
electromagnetic induction 81
electromagnetic radiation 175
electromagnetism 58, 109, 174, 215
electromotive force 7
e.m.f. 28
emission spectra 25, 26
emitter 126
emitter follower 130
energy bands 109
energy levels 109
energy spectrum 191
energy transfer in an electric circuit 18
environmental physics 215, 270
environmental pollution 249
epicentral distance 246
equivalent internal resistances 10
equivalent resistance 8
extrinsic semiconductors 114

F

Faraday's dark space 22
Faraday's law 82
Fermi energy 112
Fermi level 112
ferromagnetic materials 92, 96
fill factor 232
fission chain 206
Fleming's left-hand 58
flow rate of water 215
fluorescent lighting 26
flux density 98
forbidden energy gap 111
forward biased 117
Fourier's law 225
free electrons 2
full-wave rectification 121

G

galvanometer 14
gas laser 196
geomagnetic field 100, 102
geothermal energy 236
geothermal heat pumps 236
germanium 112
Global Positioning System 100
granular 220
green-house gases 217

H

half cycle 30
half-life 203
half- wave rectification 121
helical path 62
Henry 86
hologram 197
humidity 218
hydraulic conductivity 224
hydrocarbons 250
hydrogen spectrum 191
hysteresis 97
hysteresis loop 98

I

impedance 40
induced e.m.f. 35, 83
inductance 41
inductive reactance 36
inductor 35
Inductor capacitor (LC) circuit 49
Inductor resistor (LR) circuit 42, 48
input characteristics 132
instantaneous current 1
instantaneous phase 29
instantaneous power 35
instantaneous voltage 36
insulator 112
internal photoelectric effect 120
internal resistance 7
intrinsic semiconductors 114
ionization curve 24
ionization energy 190
ionization processes 23
ionized gases 23
isobaric 202
isobars 199
isotopes 199

J

junction 10

K

Kirchhoff's current law 11
Kirchhoff's law 10, 217
Kirchhoff's Voltage Law 11
knee voltage 117

L

Laser 195
laws of Boolean algebra 143
leakage current 133
Lenz's law 83
light-emitting diodes 120
light intensity 176
linear absorbers 239
line spectra 25
liquid laser 197
literal 142
load line 137
loam sand soil 224
logic gates 142
Lorentz force 58
love waves 244

M

magma 236
magnetic field intensity 94
magnetic field of the earth 100
magnetic field reversal 102
magnetic fields 58
magnetic field strength 63
magnetic flux 81
magnetic flux density 58
magnetic force 58

magnetic properties 92
magnetic susceptibility 93
magnetization 93
manometers 224
mass defect 199
massive 221
mass number 198
mass spectrometer 69
mean and root mean square (rms) 29
metastable state 196
micro-organisms 219
monochromatic 196
monochromatic light 176
moving coil galvanometer 68
mulching 226
mutual inductance 88
mutual-induction 85

N

nitrogen dioxide 250
NOT gate 144
npn transistor 125
n-type semiconductors 115
nuclear 23
nuclear mass 199
nuclear physics 198
nuclear power plant 208
nuclear radiations 204
nuclear reactor 207
nuclear stability 201
nuclear structure 198
nuclear waste 252

O

Oersted's experiment 64
Ohm's law 15
operational amplifier 130, 150, 151, 154, 155, 159, 172, 274
optical spectra of gases 25
output characteristics 133
overtopping devices 239

P

paleomagnetism 102
parallel combination 8
parallel connection of the cells 10
paramagnetic 92
particulate matter 249
peak current 36
peak value of current 29
pectrometer 26
pentavalent 114
permeability 224
permeability of free 74
phase angle 42
photo-diode 119
photoelectric current 177
photoelectric effect 174
photoelectric emission 177
photoelectrons 175
photosynthesis 218
Planck's blackbody radiation 175
Planck's quantum theory 178
planetary model 188
plum pudding 187
p-n junction 116
pnp transistor 125
point absorber 239
polar wandering 102
pollution 249
population inversion 195, 196
potential barrier 116
potential drop 7
potential difference (p.d.) 2,3,7,8,9,13,14,16,17, 18, 21, 23, 24,25, 37,39, 40,41, 44, 46, 49,50, 53,70, 71, 103, 104,111,119, 127,151,153, 159, 183, 212,265, 268,274
potentiometer 16
power coefficient 234
power factor 42, 46
power gain 128
pressure 21
primary waves 243
principal quantum number 192
principle of potentiometer 17
prismatic 221
p-type semiconductors 115

Q

Q-point 137
quality factor 45
quantum mechanics 174
quiescent state 135

R

radioactive decay 203
radioactive elements 201
radioactivity 198, 202
radioisotopes 204
Rayleigh–Jeans law 174
Rayleigh waves 244
reciprocity theorem 87
relative humidity 218
relaxation time 5
remanence 98
renewable energy 228, 238, 256, 269
resistance 4-20, 30-32, 35, 41-43, 45, 46, 52-57, 95, 106, 107, 120, 124, 125, 127-133, 135, 136, 137, 139, 142, 232, 233, 265
resistivity 4
resistivity of the material 19
resistor- capacitor (RC) circuit 40
resistor inductor capacitor (R-L-C) circuit 43, 50
resistors and cells 7
resonance peak 46
retentivity 98
reverse biased. 117
right-hand grip rule 64
root mean square (rms) 31
rotating coil 83
rules of boolean algebra 143
Rutherford planetary model 187
Rydberg constant 188, 191

S

saturated vapor pressure 218
saturation 136
secondary waves 244
secular variations 102
seismic waves 243
seismograms 245
seismometer 245
self-induced e.m.f. 88
self-inductance 35, 88
semiconductor lasers 197
semiconductors 112
series combination 7
series connection 9

shading 227
silicon 112
single grained 221
single stage (c-e) amplifier 134
sinusoidal current 40
soil 219
soil components 219
soil properties 220
soil structure 220
soil temperature 225, 227
soil texture 222, 269
soil-water potential 224
solar energy 229
solar irradiance 217
solar photovoltaics 229
solar radiation 215
solar spectrum 216
solenoid 79
somatic effects 252
spectrometer 26
stimulated emission 196
stopping potential 177
sulphur dioxide 250
superconductor 113

T

telecommunication 109, 160, 267, 275
temperature coefficient of the resistance 5, 6
terminators 239
terrestrial 216
terrestrial radiation 216
tesla 59
textural triangle method 222
thermal energy 2
thermal neutrons 206
thermal runway 134
threshold frequency 177, 179
threshold wavelength 180
tidal barrage system 237
tidal energy 237
tidal power 237
tidal stream systems 237
toroid 80
torque 66
transfer characteristics 133
transformer 98
transducer 160, 161
transistor as a switch 138
transistors 125
truth table 142
types of laser 196

U

ultraviolet catastrophe 174

V

valence band 110
vapour lamps 27
vapour pressure 218
variable 142
virtual value 32
visible radiation 215
vitrification process 253
voltage 29
voltage divider 17
voltage gain 128

W

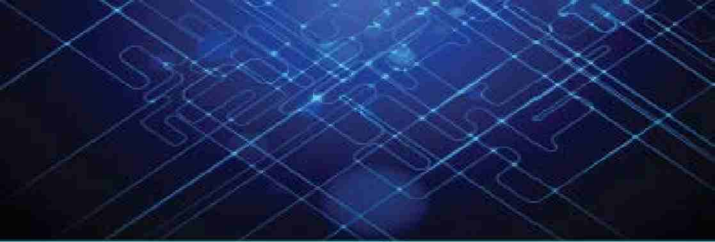
water wave 238
wave energy 238
wavelength 26
wave-particle duality 182
Wheatstone bridge 13
Wheatstone bridge circuit 13
Wheatstone meter bridge 14
Wien's displacement law 216
wind 219
wind break 226, 227
wind energy 233
wind turbine 235
work function 179

X

X-rays 184
X-rays and ultraviolet light 23

Z

Zener breakdown 118
Zener diode 118, 119.



ISBN 978-9987-09-027-3

9 789987 090273